

(19)



(11)

EP 1 592 789 B1

(12)

EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention
of the grant of the patent:

20.05.2009 Bulletin 2009/21

(51) Int Cl.:

C12N 15/00 ^(2006.01)

C12N 15/63 ^(2006.01)

C12P 21/00 ^(2006.01)

A01K 67/027 ^(2006.01)

(21) Application number: **03800225.9**

(86) International application number:

PCT/US2003/041335

(22) Date of filing: **24.12.2003**

(87) International publication number:

WO 2004/067707 (12.08.2004 Gazette 2004/33)

(54) **ADMINISTRATION OF TRANSPOSON-BASED VECTORS TO REPRODUCTIVE ORGANS**

ZUFUHR TRANSPOSONBASIERTER VEKTOREN ZU FORTPFLANZUNGSORGANEN

ADMINISTRATION DE VECTEURS A BASE DE TRANSPOSON A DES ORGANES
REPRODUCTEURS

(84) Designated Contracting States:

**AT BE BG CH CY CZ DE DK EE ES FI FR GB GR
HU IE IT LI LU MC NL PT RO SE SI SK TR**

Designated Extension States:

AL LT LV MK

(74) Representative: **Dey, Michael et al**

Weickmann & Weickmann

Patentanwälte

Postfach 86 08 20

81635 München (DE)

(30) Priority: **21.01.2003 US 441377 P**

21.01.2003 US 441381 P

21.01.2003 US 441392 P

21.01.2003 US 441405 P

21.01.2003 US 441447 P

21.01.2003 US 441502 P

26.06.2003 US 609019

(56) References cited:

WO-A-01/71019

WO-A1-01/71019

- **M. VON SPECHT; DISSERTATION: "Expression eines rekombinanten humanen Proteins in vitro und in vivo in Eileiterzellen des Huhnes, am Beispiel von humanem Erythropoietin, hrEPO" 2002, GENZENTRUM DER LUDWIG-MAXIMILIAN-UNIVERSITÄT MÜNCHEN, MÜNCHEN, XP002371226 see pages 49-68**

- **SHERMAN A ET AL: "Transposition of the Drosophila element mariner into the chicken germ line" NATURE BIOTECHNOLOGY, NATURE PUBLISHING, US, vol. 16, November 1998 (1998-11), pages 1050-1053, XP002086540 ISSN: 1087-0156**

- **OCHIAI H ET AL: "SYNTHESIS OF HUMAN ERYTHROPOIETIN IN VIVO IN THE OVIDUCT OF LAYING HENS BY LOCALIZED IN VIVO GENE TRANSFER USING ELECTROPORATION" POULTRY SCIENCE, CHAMPAIGN, IL, US, vol. 77, no. 2, 1998, pages 299-302, XP000863530 ISSN: 0032-5791**

- **PAIN B ET AL: "CHICKEN EMBRYONIC STEM CELLS AND TRANSGENIC STRATEGIES" CELLS TISSUES ORGANS, KARGER, BASEL, CH, vol. 165, 1999, pages 212-219, XP000882175 ISSN: 1422-6405**

(43) Date of publication of application:

09.11.2005 Bulletin 2005/45

(73) Proprietors:

- **TransGenRx, Inc.**

Baton Rouge, LA 70894 (US)

- **THE BOARD OF SUPERVISORS OF LOUISIANA STATE**

UNIVERSITY AND AGRICULTURAL AND MECHANICAL COLLEGE

Baton Rouge, LA 70894 (US)

(72) Inventors:

- **COOPER, Richard, K.**

Baton Rouge, LA 70810 (US)

- **FIORETTI, William, C.**

Addison, TX 75001 (US)

- **CADD, Gary G.**

Grapevine, TX 76051 (US)

Note: Within nine months of the publication of the mention of the grant of the European patent in the European Patent Bulletin, any person may give notice to the European Patent Office of opposition to that patent, in accordance with the Implementing Regulations. Notice of opposition shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

EP 1 592 789 B1

- SARMASIK ALIYE ET AL: "Transgenic live-bearing fish and crustaceans produced by transforming immature gonads with replication-defective pantropic retroviral vectors" MARINE BIOTECHNOLOGY (NEW YORK), vol. 3, no. 5, September 2001 (2001-09), pages 470-477, XP002371224 ISSN: 1436-2228
- MOHAMMED S M ET AL: "Deposition of genetically engineered human antibodies into the egg yolk of hens" IMMUNOTECHNOLOGY, ELSEVIER SCIENCE PUBLISHERS BV, NL, vol. 4, no. 2, October 1998 (1998-10), pages 115-125, XP004153636 ISSN: 1380-2933
- HORN. C. AND ERNST A. WIMMER: 'A versatile vector set for animal transgenesis' DEVELOPMENT GENES AND EVOLUTION vol. 210, no. 12, 2000, pages 630 - 637, XP002981874
- EGGLESTON P. AND YUGUANG ZHAO: 'A sensitive and rapid assay for homologous recombination in mosquito cells: impact of vector topology and implications for gene targeting' BMC GENETICS vol. 2, no. 21, 17 December 2001, pages 1 - 9, XP002981875

Description

FIELD OF THE INVENTION

5 [0001] The present invention relates to a method of producing proteins, polypeptides or peptides comprising administering a composition comprising a transposon-based vector to an oviduct or an ovary of an animal. Such administration results in incorporation of a gene of interest contained in the vector in the ovary, the oviduct or an ovum of the animal. The present invention further includes production of a protein encoded by the gene in an egg produced by the animal.

10 BACKGROUND OF THE INVENTION

[0002] Transgenic animals are desirable for a variety of reasons, including their potential as biological factories to produce desired molecules for pharmaceutical, diagnostic and industrial uses. This potential is attractive to the industry due to the inadequate capacity in facilities used for recombinant production of desired molecules and the increasing demand by the pharmaceutical industry for use of these facilities. Numerous attempts to produce transgenic animals have met several problems, including low rates of gene incorporation and unstable gene incorporation. Accordingly, improved gene technologies are needed for the development of transgenic animals for the production of desired molecules.

20 [0003] Improved gene delivery technologies are also needed for the treatment of disease in animals and humans. Many diseases and conditions can be treated with gene-delivery technologies, which provide a gene of interest to a patient suffering from the disease or the condition. An example of such disease is Type 1 diabetes. Type 1 diabetes is an autoimmune disease that ultimately results in destruction of the insulin producing β -cells in the pancreas. Although patients with Type 1 diabetes may be treated adequately with insulin injections or insulin pumps, these therapies are only partially effective. Insulin replacement, such as via insulin injection or pump administration, cannot fully reverse the defect in the vascular endothelium found in the hyperglycemic state (Pieper et al., 1996. Diabetes Res. Clin. Pract. Suppl. S157-S162). In addition, hyper- and hypoglycemia occurs frequently despite intensive home blood glucose monitoring. Finally, careful dietary constraints are needed to maintain an adequate ratio of calories consumed. This often causes major psychosocial stress for many diabetic patients. Development of gene therapies providing delivery of the insulin gene into the pancreas of diabetic patients could overcome many of these problems and result in improved life expectancy and quality of life.

[0004] Several of the prior art gene delivery technologies employed viruses that are associated with potentially undesirable side effects and safety concerns. The majority of current gene-delivery technologies useful for gene therapy reply on virus-based delivery vectors, such as adeno and adeno-associated viruses, retroviruses, and other viruses, which have been attenuated to no longer replicate. (Kay, M.A., et al. 2001. Nature Medicine 7:33-40).

35 [0005] There are multiple problems associated with the use of viral vectors. Firstly, they are not tissue-specific. In fact, a gene therapy trial using adenovirus was recently halted because the vector was present in the patient's sperm (Gene trial to proceed despite fears that therapy could change child's genetic makeup. The New York Times, December 23, 2001). Secondly, viral vectors are likely to be transiently incorporated, which necessitates re-treating a patient at specified time intervals. (Kay, M.A., et al. 2001. Nature Medicine 7:33-40). Thirdly, there is a concern that a viral-based vector could revert to its virulent form and cause disease. Fourthly, viral-based vectors require a dividing cell for stable integration. Fifthly, viral-based vectors indiscriminately integrate into various cells, which can result in undesirable germline integration. Sixthly, the required high titers needed to achieve the desired effect have resulted in the death of one patient and they are believed to be responsible for induction of cancer in a separate study. (Science, News of the Week, October 4, 2002).

45 [0006] Accordingly, what is needed is a new method to produce transgenic animals and humans with stably incorporated genes, in which the vector containing those genes does not cause disease or other unwanted side effects. There is also a need for DNA constructs that would be stably incorporated into the tissues and cells of animals and humans, including cells in the resting state that are not replicating. There is a further recognized need in the art for DNA constructs capable of delivering genes to specific tissues and cells of animals and humans.

50 [0007] When incorporating a gene of interest into an animal for the production of a desired protein or when incorporating a gene of interest in an animal or human for the treatment of a disease, it is often desirable to selectively activate incorporated genes using inducible promoters. These inducible promoters are regulated by substances either produced or recognized by the transcription control elements within the cell in which the gene is incorporated. In many instances, control of gene expression is desired in transgenic animals or humans so that incorporated genes are selectively activated at desired times and/or under the influence of specific substances. Accordingly, what is needed is a means to selectively activate genes introduced into the genome of cells of a transgenic animal or human. This can be taken a step further to cause incorporation to be tissue-specific, which prevents wide-spread gene incorporation throughout a patient's body (animal or human). This decreases the amount of DNA needed for a treatment, decreases the chance of incorporation

in gametes, and targets gene delivery, incorporation, and expression to the desired tissue where the gene is needed to function. What is also needed is a rapid expression method for rapidly producing a protein or peptide of interest in eggs and milk of transgenic animals.

5 SUMMARY OF THE INVENTION

[0008] The present invention addresses the problems described above by providing a method according to claim 1.

[0009] Animals are made transgenic through administration of a composition comprising a transposon-based vector designed for incorporation of a gene of interest for production of a desired protein, together with an acceptable carrier.

10 The compositions used according to the present invention are introduced into an oviduct or an ovary of a bird. The compositions used according to the present invention may be administered to a reproductive organ of an animal through the cloaca. The compositions used according to the present invention may be directly administered to an oviduct or an ovary or can be administered to an artery leading to an oviduct or an ovary. A transfection reagent is optionally added to the composition before administration.

15 **[0010]** The transposon-based vectors of the present invention include a transposase, operably-linked to a first promoter, and a coding sequence for a protein or peptide of interest operably-linked to a second promoter, wherein the coding sequence for the protein or peptide of interest and its operably-linked promoter are flanked by transposase insertion sequences recognized by the transposase and wherein the first promoter comprises a modified Kozak sequence comprising ACC ATG (SEQ ID NO:1). The transposon-based vector also includes the following characteristics: a) one or
20 more modified Kozak sequences at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino acids of the transposase, wherein the nucleotide at the third base position of each codon is changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) addition of an effective polyA sequence operably linked to the transposase to further enhance expression of the transposase gene. In some
25 embodiments, the effective polyA sequence is an avian optimized polyA sequence.

[0011] The present invention also provides for tissue-specific incorporation and/or expression of a gene of interest. Tissue-specific incorporation of a gene of interest, may be achieved by placing the transposase gene under the control of a tissue-specific promoter, whereas tissue-specific expression of a gene of interest may be achieved by placing the
30 gene of interest under the control of a tissue-specific promoter. In some embodiments, the gene of interest is transcribed under the influence of an ovalbumin, or other oviduct specific, promoter. Linking the gene of interest to an oviduct specific promoter in an egg-laying animal results in synthesis of a desired molecule and deposition of the desired molecule in a developing egg.

[0012] The present invention advantageously produces a high number of transgenic animals having a gene of interest stably incorporated. In some embodiments wherein the transposon-based vector is administered to the ovary, these
35 transgenic animals successfully pass the desired gene to their progeny. Accordingly, the present invention can be used to obtain transgenic animals having the gene of interest incorporated into the germline through transfection of the ovary or the present invention can be used to obtain transgenic animals having the gene of interest incorporated into the oviduct in a tissue-specific manner. Both types of transgenic animals of the present invention produce large amounts of a desired molecule encoded by the transgene. Transgenic egg-laying animals, particularly avians, produce large amounts
40 of a desired protein that is deposited in the egg for rapid harvest and purification.

[0013] Any desired gene may be incorporated into the novel transposon-based vectors of the present invention in order to synthesize a desired molecule in the transgenic animals. Proteins, peptides and nucleic acids are the desired
45 molecules to be produced by the transgenic animals of the present invention. Particularly preferred proteins are antibody proteins and other immunopharmaceutical proteins.

[0014] This invention provides the use of a composition useful for the production of transgenic hens capable of producing substantially high amounts of a desired protein or peptide. Entire flocks of transgenic birds may be developed
50 very, quickly in order to produce industrial amounts of desired molecules. The present invention solves the problems inherent in the inadequate capacity of fermentation facilities used for bacterial production of molecules and provides a more efficient and economical way to produce desired molecules. Accordingly, the present invention provides a means to produce large amounts of therapeutic, diagnostic and reagent molecules.

[0015] Transgenic chickens are excellent in terms of convenience and efficiency of manufacturing molecules such as proteins and peptides. Starting with a single transgenic rooster, thousands of transgenic offspring can be produced
55 within a year. (In principle, up to forty million offspring could be produced in just three generations). Each transgenic female is expected to lay at least 250 eggs/year, each potentially containing hundreds of milligrams of the selected protein. Flocks of chickens numbering in the hundreds of thousands are readily handled through established commercial systems. The technologies for obtaining eggs and fractionating them are also well known and widely accepted. Thus, for each therapeutic, diagnostic, or other protein of interest, large amounts of a substantially pure material can be produced at relatively low incremental cost.

[0016] A wide range of recombinant peptides and proteins can be produced in transgenic egg-laying animals. Enzymes, hormones, antibodies, growth factors, serum proteins, commodity proteins, biological response modifiers, peptides and designed proteins may all be made through practice of the present invention. For example, rough estimates suggest that it is possible to produce in bulk growth hormone, insulin, or Factor VIII, and deposit them in egg whites, for an incremental cost in the order of one dollar per gram. At such prices it is feasible to consider administering such medical agents by inhalation or even orally, instead of through injection. Even if bioavailability rates through these avenues were low, the cost of a much higher effective-dose would not be prohibitive.

[0017] In one embodiment, the egg-laying transgenic animal is an avian. The methods of the present invention may be used in avians including Ratites, Psittaciformes, Falconiformes, Piciformes, Strigiformes, Passeriformes, Coraciiformes, Ralliformes, Cuculiformes, Columbiformes, Galliformes, Anseriformes, and Herodiones. Preferably, the egg-laying transgenic animal is a poultry bird. More preferably, the bird is a chicken, turkey, duck, goose or quail. Another preferred bird is a ratite, such as, an emu, an ostrich, a rhea, or a cassowary. Other preferred birds are partridge, pheasant, kiwi, parrot, parakeet, macaw, falcon, eagle, hawk, pigeon, cockatoo, song birds, jay bird, blackbird, finch, warbler, canary, toucan, mynah, or sparrow.

[0018] The present invention makes reference to novel transposon-based vectors.

[0019] The present invention makes reference to novel transposon-based vectors that encode for the production of desired proteins or peptides in cells.

The present invention makes reference to the production of transgenic animals through intraoviduct or intraovarian administration of a transposon-based vector.

The present invention makes reference to the production of transgenic animals through intraoviduct or intraovarian administration of a transposon-based vector, wherein the transgenic animals produce desired proteins or peptides.

[0020] The present invention makes reference to a method to produce transgenic animals through intraovarian administration of a transposon-based vector that are capable of producing transgenic progeny.

[0021] An object of the present invention is to provide a method to produce transgenic animals through intraoviduct or intraovarian administration of a transposon-based vector that are capable of producing a desired molecule, such as a protein, peptide or nucleic acid.

[0022] Another object of the present invention is to provide a method to produce transgenic animals through intraoviduct or intraovarian administration of a transposon-based vector, wherein such administration results in modulation of endogenous gene expression.

[0023] It is yet another object of the present invention to provide a method to produce transgenic avians through intraoviduct or intraovarian administration of a transposon-based vector that are capable of producing proteins, peptides or nucleic acids.

[0024] It is another object of the present invention to produce transgenic animals through intraoviduct or intraovarian administration of a transposon-based vector encoding an antibody or a fragment thereof.

[0025] Still another object of the present invention is to provide a method to produce transgenic avians through intraoviduct or intraovarian administration of a transposon-based vector that are capable of producing proteins or peptides and depositing these proteins or peptides in the egg.

[0026] Another object of the present invention is to provide transgenic avians that contain a stably incorporated transgene.

[0027] Still another object of the present invention is to provide eggs containing desired proteins or peptides encoded by a transgene incorporated into the transgenic avian that produces the egg.

An advantage of the present invention is that transgenic animals are produced by the method of the present invention with higher efficiencies than observed in the prior art.

[0028] Another advantage of the present invention is that these transgenic animals possess high copy numbers of the transgene.

[0029] Another advantage of the present invention is that the transgenic animals produce large amounts of desired molecules encoded by the transgene.

[0030] Still another advantage of the present invention is that desired molecules are produced by the transgenic animals much more efficiently and economically than prior art methods, thereby providing a means for large scale production of desired molecules, particularly proteins and peptides.

[0031] Yet another advantage of the present invention is that the desired proteins and peptides are produced rapidly after making animals transgenic through introduction of the vectors of the present invention.

[0032] These and other objects, features and advantages of the present invention will become apparent after a review of the following detailed description of the disclosed embodiments and claims.

BRIEF DESCRIPTION OF THE FIGURES

[0033]

Figure 1 depicts schematically a transposon-based vector containing a transposase operably linked to a first promoter and a gene of interest operably-linked to a second promoter, wherein the gene of interest and its operably-linked promoter are flanked by insertion sequences (IS) recognized by the transposase. "Pro" designate a promoter. In this and subsequent figures, the size of the actual nucleotide sequence is not necessarily proportionate to the box representing that sequence.

Figure 2 depicts schematically a transposon-based vector for targeting deposition of a polypeptide in an egg white wherein Ov pro is the ovalbumin promoter, Ov protein is the ovalbumin protein and PolyA is a polyadenylation sequence. The TAG sequence includes a spacer sequence, the gp41 hairpin loop from HIV I and a protease cleavage site.

Figure 3 depicts schematically a transposon-based vector for targeting deposition of a polypeptide in an egg white wherein Ovo pro is the ovomucoid promoter and Ovo SS is the ovomucoid signal sequence. The TAG sequence includes a spacer, the gp41 hairpin loop from HIV I and a protease cleavage site.

Figure 4 depicts schematically a transposon based-vector for expression of an RNAi molecule. "Tet pro" indicates a tetracycline inducible promoter whereas "pro" indicates the pro portion of a prepro sequence as described herein "Ovgen" indicates approximately 60 base pairs of an ovalbumin gene, "Ovotraas" indicates approximately 60 base pairs of an ovotransferrin gene and "Ovomucin" indicates approximately 60 base pairs of an ovomucin gene.

Figure 5 is a picture of an SDS-PAGE gel wherein a pooled fraction of an isolated proinsulin fusion protein was run in lanes 4 and 6. Lanes 1 and 10 of the gel contain molecular weight standards, lanes 2 and 8 contain non-transgenic chicken egg white, and lanes 3, 5, 7 and 9 are blank.

DETAILED DESCRIPTION OF THE INVENTION

[0034] The present invention provides a new, effective and efficient method of producing transgenic animals, i.e. birds, through administration of a composition comprising a transposon-based vector designed for incorporation of a gene of interest and production of a desired molecule. The transposon-based vectors are administered to an oviduct or an ovary.

[0035] The vectors may be directly administered to an oviduct or an ovary or can be administered to an artery leading to an oviduct or an ovary or to a lymph system proximate to the cells to be genetically altered. The vectors may be administered to an oviduct or an ovary of an animal through the cloaca. One method of direct administration is by injection, and in one embodiment, the lumen of the magnum of the oviduct is injected with a transposon-based vector. Another method of direct administration is by injection, and in one embodiment, the lumen of, the infundibulum of the oviduct is injected with a transposon-based vector. A preferred intrarterial administration is an administration into an artery that supplies the oviduct or the ovary. In some embodiments, administration of the transposon-based vector to an oviduct or an artery that leads to the oviduct results in incorporation of the vector into the epithelial and/or secretory cells of the oviduct. In other embodiments, administration of the transposon-based vector to an ovary or an artery that leads to the ovary or a lymphatic system proximal to the ovary results in incorporation of the vector into an oocyte or a germinal disk inside the ovary.

Definition

[0036] It is to be understood that as used in the specification and in the claims, "a" or "an" can mean one or more, depending upon the context in which it is used. Thus, for example, reference to "a cell" can mean that at least one cell can be utilized.

[0037] The term "antibody" is used interchangeably with the term "immunoglobulin" and is defined herein as a protein synthesized by an animal or a cell of the immune system in response to the presence of a foreign substance commonly referred to as an "antigen" or an "immunogen". The term antibody includes fragments of antibodies. Antibodies are characterized by specific affinity to a site on the antigen, wherein the site is referred to an "antigenic determinant" or an "epitope". Antigens can be naturally occurring or artificially engineered. Artificially engineered antigens include, but are not limited to, small molecules, such as small peptides, attached to haptens such as macromolecules, for example proteins, nucleic acids, or polysaccharides. Artificially designed or engineered variants of naturally occurring antibodies and artificially designed or engineered antibodies not occurring in nature are all included in the current definition. Such variants include conservatively substituted amino acids and other forms of substitution as described in the section concerning proteins and polypeptides.

[0038] As used herein, the term "egg-laying animal" includes all amniotes such as birds, turtles, lizards and monotremes. Monotremes are egg-laying mammals and include the platypus and echidna. The term "bird" or "fowl," as used herein, is defined as a member of the Aves class of animals which are characterized as warm-blooded, egg-laying vertebrates primarily adapted for flying. Avians include, without limitation, Ratites, Psittaciformes, Falconiformes, Piciformes, Strigiformes, Passeriformes, Coraciiformes, Ralliformes, Cuculiformes, Columbiformes, Galliformes, Anseriformes, and Herodiones. The term "Ratite," as used herein, is defined as a group of flightless, mostly large, running

birds comprising several orders and including the emus, ostriches, kiwis, and cassowaries. The term "Psittaciformes", as used herein, includes parrots and refers to a monofamilial order of birds that exhibit zygodactylism and have a strong hooked bill. A "parrot" is defined as any member of the avian family Psittacidae (the single family of the Psittaciformes), distinguished by the short, stout, strongly hooked beak. Avians include all poultry birds, especially chickens, geese, turkeys, ducks and quail. The term "chicken" as used herein denotes chickens used for table egg production, such as egg-type chickens, chickens reared for public meat consumption, or broilers, and chickens reared for both egg and meat production ("dual-purpose" chickens). The term "chicken," also denotes chickens produced by primary breeder companies, or chickens that are the parents, grandparents, great-grandparents, etc. of those chickens reared for public table egg, meat, or table egg and meat consumption.

[0039] The term "egg" is defined herein as including a large female sex cell enclosed in a porous, calcareous or leathery shell, produced by birds and reptiles. The term "ovum" is defined as a female gamete, and is also known as an egg. Therefore, egg production in all animals other than birds and reptiles, as used herein, is defined as the production and discharge of an ovum from an ovary, or "ovulation". Accordingly, it is to be understood that the term "egg" as used herein is defined as a large female sex cell enclosed in a porous, calcareous or leathery shell, when a bird or reptile produces it, or it is an ovum when it is produced by all other animals.

[0040] The term "milk-producing animal" refers herein to mammals including, but not limited to, bovine, ovine, porcine, equine, and primate animals. Milk-producing animals include but are not limited to cows, llamas, camels, goats, reindeer, zebu, water buffalo, yak, horses, pigs, rabbits, non-human primates, and humane.

[0041] The term "gene" is defined herein to include a coding region for a protein, peptide or polypeptide.

[0042] The term "transgenic animal" refers to an animal having at least a portion of the transposon-based vector DNA incorporated into its DNA. While a transgenic animal includes an animal wherein the transposon-based vector DNA is incorporated into the germline DNA, a transgenic animal also includes an animal having DNA in one or more cells that contain a portion of the transposon-based vector DNA for any period of time. In a preferred embodiment, a portion of the transposon-based vector comprises a gene of interest. More preferably, the gene of interest is incorporated into the animal's DNA for a period of at least five days, more preferably the reproductive life of the animal, and most preferably the life of the animal. In a further preferred embodiment, the animal is an avian.

[0043] The term "vector" is used interchangeably with the terms "construct", "DNA construct" and "genetic construct" to denote synthetic nucleotide sequences used for manipulation of genetic material, including but not limited to cloning, subcloning, sequencing, or introduction of exogenous genetic material into cells, tissues or organisms, such as birds. It is understood by one skilled in the art that vectors may contain synthetic DNA sequences, naturally occurring DNA sequences, or both. The vectors of the present invention are transposon-based vectors as described herein.

[0044] When referring to two nucleotide sequences, one being a regulatory sequence, the term "operably-linked" is defined herein to mean that the two sequences are associated in a manner that allows the regulatory sequence to affect expression of the other nucleotide sequence. It is not required that the operably-linked sequences be directly adjacent to one another with no intervening sequence(s).

[0045] The term "regulator sequence" is defined herein as including promoters, enhancers and other expression control elements such as polyadenylation sequences, matrix attachment sites, insulator regions for expression of multiple genes on a single construct, ribosome entry/attachment sites, introns that are able to enhance expression, and silencers.

Transposon-Based Vectors

[0046] Transposon-based vectors according to the invention are transposon-based vectors which are used in the method of the present invention.

[0047] While not wanting to be bound by the following statement, it is believed that the nature of the DNA construct is an important factor in successfully producing transgenic animals. The "standard" types of plasmid and viral vectors that have previously been almost universally used for transgenic work in all species, especially avians, have low efficiencies and may constitute a major reason for the low rates of transformation previously observed. The DNA (or RNA) constructs previously used often do not integrate into the host DNA, or integrate only at low frequencies. Other factors may have also played a part, such as poor entry of the vector into target cells. The present invention provides transposon-based vectors that can be administered to an animal that overcome the prior art problems relating to low transgene integration frequencies. Two preferred transposon-based vectors of the present invention in which a transposase, gene of interest and other polynucleotide sequences may be introduced are termed pTnMCS (SEQ ID NO:2) and pTnMod (SEQ ID NO:3).

[0048] The transposon-based vectors of the present invention produce integration frequencies an order of magnitude greater than has been achieved with previous vectors. More specifically, intratesticular injections performed with a prior art transposon-based vector (described in U.S. Patent No. 5,719,055) resulted in 41% sperm positive roosters whereas intratesticular injections performed with the novel transposon-based vectors of the present invention resulted in 77% sperm positive roosters. Actual frequencies of integration were estimated by either or both comparative strength of the

PCR signal from the sperm and histological evaluation of the testes and sperm by quantitative PCR.

[0049] The transposon-based vectors of the present invention include a transposase gene operably-linked to a first promoter, and a coding sequence for a desired protein or peptide operably-linked to a second promoter, wherein the coding sequence for the desired protein or peptide and its operably-linked promoter are flanked by transposase insertion sequences recognized by the transposase. The transposon-based vector also includes one or more of the following characteristics: a) one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:1) at the 3' end of the first promoter to enhance expression of the transposase; b) modifications of the codons for the first several N-terminal amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene. The transposon-based vector may additionally or alternatively include one or more of the following Kozak sequences at the 3' end of any promoter, including the promoter operably-linked to the transposase: ACCATGG (SEQ ID NO:4), AAGATGT (SEQ ID NO:5), ACGATGA (SEQ ID NO:6), AAGATGG (SEQ ID NO:7), GACATGA (SEQ ID NO:8), ACCATGA (SEQ ID NO:9), and ACCATGA (SEQ ID NO:10), ACCATGT (SEQ ID NO:52).

[0050] Figure 1 shows a schematic representation of several components of the transposon-based vector. The present invention further includes vectors containing more than one gene of interest, wherein a second or subsequent gene of interest is operably-linked to the second promoter or to a different promoter. It is also to be understood that the transposon-based vectors shown in the Figures are representative of the present invention and that the order of the vector elements may be different than that shown in the Figures, that the elements may be present in various orientations, and that the vectors may contain additional elements not shown in the Figures.

Transposases and Insertion Sequences

[0051] In a further embodiment of the present invention, the transposase found in the transposase-based vector is an altered target site (ATS) transposase and the insertion sequences are those recognized by the ATS transposase. However, the transposase located in the transposase-based vectors is not limited to a modified ATS transposase and can be derived from any transposase. Transposases known in the prior art include those found in AC7, Tn5SEQ1, Tn916, Tn951, Tn1721, Tn 2410, Tn1681, Tn1, Tn2, Tn3, Tn4, Tn5, Tn6, Tn9, Tn10, Tn30, Tn101, Tn903, Tn501, Tn1000 ($\gamma\delta$), Tn1681, Tn2901, ACtransposons, Mptransposons, Spmtransposons, Enttransposons, Dottedtransposons, Mu transposons, Ds transposons, dSpm transposons and I transposons. According to the present invention, these transposases and their regulatory sequences are modified for improved functioning as follows: a) the addition one or more modified Kozak sequences comprising ACCATG (SEQ ID NO:1) at the 3' end of the promoter operably-linked to the transposase; b) a change of the codons for the first several amino acids of the transposase, wherein the third base of each codon was changed to an A or a T without changing the corresponding amino acid; c) the addition of one or more stop codons to enhance the termination of transposase synthesis; and/or, d) the addition of an effective polyA sequence operably-linked to the transposase to further enhance expression of the transposase gene.

[0052] Although not wanting to be bound by the following statement, it is believed that the modifications of the first several N-terminal codons of the transposase gene increase transcription of the transposase gene, in part, by increasing strand dissociation. It is preferable that between approximately 1 and 20, more preferably 3 and 15, and most preferably between 4 and 12 of the first N-terminal codons of the transposase are modified such that the third base of each codon is changed to an A or a T without changing the encoded amino acid. In one embodiment the first ten N-terminal codons of the transposase gene are modified in this manner. It is also preferred that the transposase contain mutations that make it less specific for preferred insertion sites and thus increases the rate of transgene insertion as discussed in U.S. Patent No. 5,719,055.

[0053] In some embodiments, the transposon-based vectors are optimized for expression in a particular host by changing the methylation patterns of the vector DNA. For example, prokaryotic methylation may be reduced by using a methylation deficient organism for production of the transposon-based vector. The transposon-based vectors may also be methylated to resemble eukaryotic DNA for expression in a eukaryotic host.

[0054] Transposases and insertion sequences from other analogous eukaryotic transposon-based vectors that can also be modified and used are, for example, the *Drosophila* P element derived vectors disclosed in U.S. Patent No. 6,291,243; the *Drosophila* mariner element described in Sherman et al. (1998); or the sleeping beauty transposon. See also Hackett et al. (1999); D. Lampe et al., 1999. Proc. Natl. Acad. Sci. USA, 96:11428-11433; S. Fischer et al., 2001. Proc. Natl. Acad. Sci. USA, 98:6759-6764; L. Zagoraiou et al., 2001. Proc. Natl. Acad. Sci. USA, 98:11474-11478; and D. Berg et al. (Eds.), Mobile DNA, Amer. Soc. Microbiol. (Washington, D.C., 1989). However, it should be noted that bacterial transposon-based elements are preferred, as there is less likelihood that a eukaryotic transposase in the recipient species will recognize prokaryotic insertion sequences bracketing the transgene.

[0055] Many transposases recognize different insertion sequences, and therefore, it is to be understood that a transposase-based vector will contain insertion sequences recognized by the particular transposase also found in the trans-

posase-based vector. In a preferred embodiment of the invention, the insertion sequences have been shortened to about 70 base pairs in length as compared to those found in wild-type transposons that typically contain insertion sequences of well over 100 base pairs.

[0056] While the examples provided below incorporate a "cut and insert" Tn10 based vector that is destroyed following the insertion event, the present invention also encompasses the use of a "rolling replication" type transposon-based vector. Use of a rolling replication type transposon allows multiple copies of the transposon/transgene to be made from a single transgene construct and the copies inserted. This type of transposon-based system thereby provides for insertion of multiple copies of a transgene into a single genome. A rolling replication type transposon-based vector may be preferred when the promoter operably-linked to gene of interest is endogenous to the host cell and present in a high copy number or highly expressed. However, use of a rolling replication system may require tight control to limit the insertion events to non-lethal levels. Tn1, Tn2, Tn3, Tn4, Tn5, Tn9, Tn21, Tn501, Tn551, Tn951, Tn1721, Tn2410 and Tn2603 are examples of a rolling replication type transposon, although Tn5 could be both a rolling replication and a cut and insert type transposon.

Stop Codons and PolyA Sequences.

[0057] In one embodiment, the transposon-based vector contains two stop codons operably-linked to the transposase and/or to the gene of interest. In an alternate embodiment, one stop codon of UAA or UGA is operably linked to the transposase and/or to the gene of interest.

[0058] As used herein an "effective polyA sequence" refers to either a synthetic or non-synthetic sequence that contains multiple and sequential nucleotides containing an adenine base (an A polynucleotide string) and that increases expression of the gene to which it is operably-linked. A polyA sequence may be operably-linked to any gene in the transposon-based vector including, but not limited to, a transposase gene and a gene of interest. A preferred polyA sequence is optimized for use in the host animal or human. In one embodiment, the polyA sequence is optimized for use in an avian species and more specifically, a chicken. An avian optimized polyA sequence generally contains a minimum of 40 base pairs, preferably between approximately 40 and several hundred base pairs, and more preferably approximately 75 base pairs that precede the A polynucleotide string and thereby separate the stop codon from the A polynucleotide string. In one embodiment of the present invention, the polyA sequence comprises a conalbumin polyA sequence as provided in SEQ ID NO:11 and as taken from GenBank accession # Y00407, base pairs 10651-11058. In another embodiment, the polyA sequence comprises a synthetic polynucleotide sequence shown in SEQ ID NO:12. In yet another embodiment, the polyA sequence comprises an avian optimized polyA sequence provided in SEQ ID NO: 13. A chicken optimized polyA sequence may also have a reduced amount of CT repeats as compared to a synthetic polyA sequence.

[0059] It is a surprising discovery of the present invention that such an avian optimized poly A sequence increases expression of a polynucleotide to which it is operably-linked in an avian as compared to a non-avian optimized polyA sequence. Accordingly, the present invention includes methods of or increasing incorporation of a gene of interest wherein the gene of interest resides in a transposon-based vector containing a transposase gene and wherein the transposase gene is operably linked to an avian optimized polyA sequence. The present invention also includes methods of increasing expression of a gene of interest in an avian that includes administering a gene of interest to the avian, wherein the gene of interest is operably-linked to an avian optimized polyA sequence. An avian optimized polyA nucleotide string is defined herein as a polynucleotide containing an A polynucleotide string and a minimum of 40 base pairs, preferably between approximately 40 and several hundred base pairs, and more preferably approximately 60 base pairs that precede the A polynucleotide string. The present invention further provides transposon-based vectors containing a gene of interest or transposase gene operably linked to an avian optimized polyA sequence.

Promoters and Enhancers

[0060] The first promoter operably-linked to the transposase gene and the second promoter operably-linked to the gene of interest can be a constitutive promoter or an inducible promoter. Constitutive promoters include, but are not limited to, immediate early cytomegalovirus (CMV) promoter, herpes simplex virus 1 (HSV1) immediate early promoter, SV40 promoter, lysozyme promoter, early and late CMV promoters, early and late HSV promoters, β -actin promoter, tubulin promoter, Rous-Sarcoma virus (RSV) promoter, and heat-shock protein (HSP) promoter. Inducible promoters include tissue-specific promoters, developmentally-regulated promoters and chemically inducible promoters. Examples of tissue-specific promoters include the glucose 6 phosphate (G6P) promoter, vitellogenin promoter, ovalbumin promoter, ovomucoid promoter, conalbumin promoter, ovotransferrin promoter, prolactin promoter, kidney uromodulin promoter, and placental lactogen promoter. In one embodiment, the vitellogenin promoter includes a polynucleotide sequence of SEQ ID NO: 14. The G6P promoter sequence may be deduced from a rat G6P gene untranslated upstream region provided in GenBank accession number U57552.1. Examples of developmentally-regulated promoters include the homeobox promoters and several hormone induced promoters. Examples of chemically inducible promoters include repro-

ductive hormone induced promoters and antibiotic inducible promoters such as the tetracycline inducible promoter and the zinc-inducible" metallothioneine promoter.

[0061] Other inducible promoter systems include the Lac operator repressor system inducible by IPTG (isopropyl beta-D-thiogalactoside) (Cronin, A. et al. 2001. *Genes and Development*, v. 15), ecdysone-based inducible systems (Hoppe, U. C. et al. 2000. *Mol. Ther.* 1:159-164); estrogen-based inducible systems (Brasemann, S. et al. 1993. *Proc. Natl. Acad. Sci.* 90:1657-1661); progesterone-based inducible systems using a chimeric regulator, GLVP, which is a hybrid protein consisting of the GAL4 binding domain and the herpes simplex virus transcriptional activation domain, VP16, and a truncated form of the human progesterone receptor that retains the ability to bind ligand and can be turned on by RU486 (Wang, et al. 1994. *Proc. Natl. Acad. Sci.* 91:8180-8184); CID-based inducible systems using chemical inducers of dimerization (CIDs) to regulate gene expression, such as a system wherein rapamycin induces dimerization of the cellular proteins FKBP12 and FRAP (Belshaw, P. J. et al. 1996. *J. Chem. Biol.* 3:731-738; Fan, L. et al. 1999. *Hum. Gene Ther.* 10:2273-2285; Shariat, S.F. et al. 2001. *Cancer Res.* 61:2562-2571; Spencer, D.M. 1996. *Curr. Biol.* 6: 839-847). Chemical substances that activate the chemically inducible promoters can be administered to the animal containing the transgene of interest via any method known to those of skill in the art.

[0062] Other examples of cell or tissue-specific and constitutive promoters include but are not limited to smooth-muscle SM22 promoter, including chimeric SM22alpha/telokin promoters (Hoggatt A.M. et al., 2002. *Circ Res.* 91(12):1151-9); ubiquitin C promoter (*Biochim Biophys Acta*, 2003. Jan. 3;1625(1):52-63); Hsf2 promoter; murine COMP (cartilage oligomeric matrix protein) promoter, early B cell-specific mb-1 promoter (Sigvardsson M., et al., 2002. *Mol. Cell Biol.* 22 (24):8539-51); prostate specific antigen (PSA) promoter (Yoshimura I. et al., 2002, *J. Urol.* 168(6):2659-64); exorh promoter and pineal expression-promoting element (Asaoka Y., et al., 2002. *Proc. Natl. Acad. Sci.* 99(24):15456-61); neural and liver ceramidase gene promoters (Okino N. et al., 2002. *Biochem. Biophys. Res. Commun.* 299(1):160-6); PSP94 gene promoter/enhancer (Gabril M.Y. et al., 2002. *Gene Ther.* 9(23):1589-99); promoter of the human FAT/CD36 gene (Kuriki C., et al., 2002. *Biol. Pharm. Bull.* 25(11):1476-8); VL30 promoter (Staplin W.R. et al., 2002. *Blood* October 24, 2002); and, IL-10 promoter (Brenner S., et al., 2002. *J. Biol. Chem.* December 18, 2002).

[0063] Examples of avian promoters include, but are not limited to, promoters controlling expression of egg white proteins, such as ovalbumin, ovotransferrin (conalbumin), ovomucoid, lysozyme, ovomucin, g2 ovoglobulin, g3 ovoglobulin, ovoflavoprotein, ovostatin (ovomacroglobin), cystatin, avidin, thiamine-binding protein, glutamyl aminopeptidase minor glycoprotein 1, minor glycoprotein 2; and promoters controlling expression of egg-yolk proteins, such as vitellogenin, very low-density lipoproteins, low density lipoprotein, cobalamin-binding protein, riboflavin-binding protein, biotin-binding protein (Awade, 1996. *Z. Lebensm. Unters. Forsch.* 202:1-14). An advantage of using the vitellogenin promoter is that it is active during the egg-laying stage of an animal's life-cycle, which allows for the production of the protein of interest to be temporally connected to the import of the protein of interest, into the egg yolk when the protein of interest is equipped with an appropriate targeting sequence. In some embodiments, the avian promoter is an oviduct-specific promoter. As used herein, the term "oviduct-specific promoter" includes, but is not limited to, ovalbumin; ovotransferrin (conalbumin); ovomucoid; O1, O2, O3, O4 or O5 avidin; ovomucin; g2 ovoglobulin; g3 ovoglobulin; ovoflavoprotein; and ovostatin (ovomacroglobin) promoters.

[0064] When germline transformation occurs via intraovarian administration, liver-specific promoters may be operably-linked to the gene of interest to achieve liver-specific expression of the transgene. Liver-specific promoters of the present invention include, but are not limited to, the following promoters, vitellogenin promoter, G6P promoter, cholesterol-7-alpha hydroxylase (CYP7A) promoter, phenylalanine hydroxylase (PAH) promoter, protein C gene promoter, insulin-like growth factor I (IGF-1) promoter, bilirubin UDP-glucuronosyltransferase promoter, aldolase B promoter, furin promoter, metallothioneine promoter, albumin promoter, and insulin promoter.

[0065] Also included in the present invention are promoters that can be used to target expression of a protein of interest into the milk of a milk-producing animal including, but not limited to, β lactoglobulin promoter, whey acidic protein promoter, lactalbumin promoter and casein promoter.

[0066] When germline transformation occurs via intraovarian administration, immune system-specific promoters may be operably-linked to the gene of interest to achieve immune system-specific expression of the transgene. Accordingly, promoters associated with cells of the immune system may also be used. Acute phase promoters such as interleukin (IL)-1 and IL-2 may be employed. Promoters for heavy and light chain Ig may also be employed. The promoters of the T cell receptor components CD4 and CD8, B cell promoters and the promoters of CR2 (complement receptor type 2) may also be employed. Immune system promoters are preferably used when the desired protein is an antibody protein.

[0067] Also included in this invention are modified promoters/enhancers wherein elements of a single promoter are duplicated, modified, or otherwise changed. In one embodiment, steroid hormone-binding domains of the ovalbumin promoter are moved from about -6.5 kb to within approximately the first 1000 base pairs of the gene of interest. Modifying an existing promoter with promoter/enhancer elements not found naturally in the promoter, as well as building an entirely synthetic promoter, or drawing promoter/enhancer elements from various genes together on a non-natural backbone, are all encompassed by the current invention.

[0068] Accordingly, it is to be understood that the promoters contained within the transposon-based vectors of the

present invention may be entire promoter sequences or fragments of promoter sequences. For example, in one embodiment, the promoter operably linked to a gene of interest is an approximately 900 base pair fragment of a chicken ovalbumin promoter (SEQ ID NO:15). The constitutive and inducible promoters contained within the transposon-based vectors may also be modified by the addition of one or more modified Kozak sequences of ACCATG (SEQ ID NO:1).

[0069] As indicated above, the present invention includes transposon-based vectors containing one or more enhancers. These enhancers may or may not be operably-linked to their native promoter and may be located at any distance from their operably-linked promoter. A promoter operably-linked to an enhancer and a promoter modified to eliminate repressive regulatory effects are referred to herein as an "enhanced promoter." The enhancer contained within the transposon-based vectors are preferably enhancers found in birds, and more preferably, an ovalbumin enhancer, but are not limited to these types of enhancers. In one embodiment, an approximately 675 base pair enhancer element of an ovalbumin promoter is cloned upstream of an ovalbumin promoter with 300 base pairs of spacer DNA separating the enhancer and promoter. In one embodiment, the enhancer used as a part of the present invention comprises base pairs 1-675 of a chicken ovalbumin enhancer from GenBank accession #S82527.1. The polynucleotide sequence of this enhancer is provided in SEQ ID NO:16.

[0070] Also included in some of the transposon-based vectors of the present invention are cap sites and fragments of cap sites. In one embodiment, approximately 50 base pairs of a 5' untranslated region wherein the capsite resides are added on the 3' end of an enhanced promoter or promoter. An exemplary 5' untranslated region is provided in SEQ ID NO: 17. A putative cap-site residing in this 5' untranslated region preferably comprises the polynucleotide sequence provided in SEQ ID NO:18.

[0071] In one embodiment of the present invention, the first promoter operably-linked to the transposase gene is a constitutive promoter and the second promoter operably-linked to the gene of interest is a tissue-specific promoter. In the second embodiment, use of the first constitutive promoter allows for constitutive activation of the transposase gene and incorporation of the gene of interest, into virtually all cell types, including the germline of the recipient animal. Although the gene of interest is incorporated into the germline generally, the gene of interest may only be expressed in a tissue-specific manner. A transposon-based vector having a constitutive promoter operably-linked to the transposase gene can be administered by any route; and in one embodiment, the vector is administered to an ovary, to an artery leading to the ovary or to a lymphatic system or fluid proximal to the ovary.

[0072] It should be noted that cell- or tissue-specific expression as described herein does not require a complete absence of expression in cells or tissues other than the preferred cell or tissue. Instead, "cell-specific" or "tissue-specific" expression refers to a majority of the expression of a particular gene of interest in the preferred cell or tissue, respectively.

[0073] When incorporation of the gene of interest into the germline is not preferred, the first promoter operably-linked to the transposase gene can be a tissue-specific promoter. For example, transfection of a transposon-based vector containing a transposase gene operably-linked to an oviduct specific promoter such as the ovalbumin promoter provides for activation of the transposase gene and incorporation of the gene of interest in the cells of the oviduct but not into the germline and other cells generally. In this embodiment, the second promoter operably-linked to the gene of interest can be a constitutive promoter or an inducible promoter. In a preferred embodiment, both the first promoter and the second promoter are an Ovalbumin promoter. In embodiments wherein tissue-specific expression or incorporation is desired, it is preferred that the transposon-based vector is administered directly to the tissue of interest, to an artery leading to the tissue of interest or to fluids surrounding the tissue of interest. In a preferred embodiment, the tissue of interest is the oviduct and administration is achieved by direct injection into the oviduct or an artery leading to the oviduct. In a further preferred embodiment, administration is achieved by direct injection into the lumen of the magnum or the infundibulum of the oviduct. Indirect administration to the oviduct may occur through the cloaca.

[0074] Accordingly, cell specific promoters may be used to enhance transcription in selected tissues. In birds, for example, promoters that are found in cells of the fallopian tube, such as ovalbumin, conalbumin, ovomucoid and/or lysozyme, are used in the vectors to ensure transcription of the gene of interest in the epithelial cells and tubular gland cells of the fallopian tube, leading to synthesis of the desired protein encoded by the gene and deposition into the egg white. In mammals, promoters specific for the epithelial cells of the alveoli of the mammary gland, such as prolactin, insulin, beta lactoglobulin, whey acidic protein, lactalbumin, casein, and/or placental lactogen, are used in the design of vectors used for transfection of these cells for the production of desired proteins for deposition into the milk. In liver cells, the G6P promoter may be employed to drive transcription of the gene of interest for protein production. Proteins made in the liver of birds may be delivered to the egg yolk.

[0075] In order to achieve higher or more efficient expression of the transposase gene, the promoter and other regulatory sequences operably-linked to the transposase gene may be those derived from the host. These host specific regulatory sequences can be tissue specific as described above or can be of a constitutive nature. For example, an avian actin promoter and its associated polyA sequence can be operably-linked to a transposase in a transposase-based vector for transfection into an avian. Examples of other host specific promoters that could be operably-linked to the transposase include the myosin and DNA or RNA polymerase promoters.

Directing Sequences

[0076] In some embodiments of the present invention, the gene of interest is operably-linked to a directing sequence or a sequence that provides proper conformation to the desired protein encoded by the gene of interest. As used herein, the term "directing sequence" refers to both signal sequences and targeting sequences. An egg directing sequence includes, but is not limited to, an ovomucoid signal sequence, an ovalbumin signal sequence, a cecropin pre pro signal sequence, and a vitellogenin targeting sequence. The term "signal sequence" refers to an amino acid sequence, or the polynucleotide sequence that encodes the amino acid sequence, that directs the protein to which it is linked to the endoplasmic reticulum in a eukaryote, and more preferably the translocational pores in the endoplasmic reticulum, or the plasma membrane in a prokaryote, or mitochondria, such as for the purpose of gene therapy for mitochondrial diseases. Signal and targeting sequences can be used to direct a desired protein into, for example, the milk, when the transposon-based vectors are administered to a milk-producing animal.

[0077] Signal sequences can also be used to direct a desired protein into, for example, a secretory pathway for incorporation into the egg yolk or the egg white, when the transposon-based vectors are administered to a bird or other egg-laying animal. One example of such a transposon-based vector is provided in Figure 3 wherein the gene of interest is operably linked to the ovomucoid signal sequence. The present invention also includes a gene of interest operably-linked to a second gene containing a signal sequence. An example of such an embodiment is shown in Figure 2 wherein the gene of interest is operably-linked to the ovalbumin gene that contains an ovalbumin signal sequence. Other signal sequences that can be included in the transposon-based vectors include, but are not limited to the ovotransferrin and lysozyme signal sequences. In one embodiment, the signal sequence is an ovalbumin signal sequence including a sequence shown in SEQ ID NO:19. In another embodiment, the signal sequence is a modified ovalbumin signal sequence including a sequence shown in SEQ ID NO:20 or SEQ ID NO:21.

[0078] As also used herein, the term "targeting sequence" refers to an amino acid sequence, or the polynucleotide sequence encoding the amino acid sequence, which amino acid sequence is recognized by a receptor located on the exterior of a cell. Binding of the receptor to the targeting sequence results in uptake of the protein or peptide operably-linked to the targeting sequence by the cell. One example of a targeting sequence is a vitellogenin targeting sequence that is recognized by a vitellogenin receptor (or the low density lipoprotein receptor) on the exterior of an oocyte. In one embodiment, the vitellogenin targeting sequence includes the polynucleotide sequence of SEQ ID NO:22. In another embodiment, the vitellogenin targeting sequence includes all or part of the vitellogenin gene. Other targeting sequences include VLDL and Apo E, which are also capable of binding the vitellogenin receptor. Since the ApoE protein is not endogenously expressed in birds, its presence may be used advantageously to identify birds carrying the transposon-based vectors of the present invention.

Genes of Interest Encoring Desired Proteins

[0079] A gene of interest selected for stable incorporation is designed to encode any desired protein or peptide or to regulate any cellular response. In some embodiments, the desired proteins or peptides are deposited in an egg. It is to be understood that the present invention encompasses transposon-based vectors containing multiple genes of interest. The multiple genes of interest may each be operably-linked to a separate promoter and other regulatory sequence(s) or may all be operably-linked to the same promoter and other regulatory sequence(s). In one embodiment, multiple genes of interest are linked to a single promoter and other regulatory sequence(s) and each gene of interest is separated by a cleavage site or a portion of a signal sequence. A gene of interest may contain modifications of the codons for the first several N-terminal amino acids of the gene of interest, wherein the third base of each codon is changed to an A or a T without changing the corresponding amino acid.

[0080] Protein and peptide hormones are a preferred class of proteins in the present invention. Such protein and peptide hormones are synthesized throughout the endocrine system and include, but are not limited to, hypothalamic hormones and hypophysiotropic hormones, anterior, intermediate and posterior pituitary hormones, pancreatic islet hormones, hormones made in the gastrointestinal system, renal hormones, thymic hormones, parathyroid hormones, adrenal cortical and medullary hormones. Specifically, hormones that can be produced using the present invention include, but are not limited to, chorionic gonadotropin, corticotropin, erythropoietin, glucagons, IGF-1, oxytocin, platelet-derived growth factor, calcitonin, follicle-stimulating hormone, luteinizing hormone, thyroid-stimulating hormone, insulin, gonadotropin-releasing hormone and its analogs, vasopressin, octreotide, somatostatin, prolactin, adrenocorticotrophic hormone, antidiuretic hormone, thyrotropin-releasing hormone (TRH), growth hormone-releasing hormone (GHRH), dopamine, melatonin, thyroxine (T_4), parathyroid hormone (PTH), glucocorticoids such as cortisol, mineralocorticoids such as aldosterone, androgens such as testosterone, adrenaline (epinephrine), noradrenaline (norepinephrine), estrogens such as estradiol, progesterone, glucagons, calcitonin, calciferol, atrial-natriuretic peptide, gastrin, secretin, cholecystokinin (CCK), neuropeptide Y, ghrelin, PYY₃₋₃₆, angiotensinogen, thrombopoietin, and leptin. By using appropriate polynucleotide sequences, species-specific hormones may be made by transgenic animals.

[0081] In one embodiment of the present invention, the gene of interest is a proinsulin gene and the desired molecule is insulin. Proinsulin consists of three parts: a C-peptide and two strands of amino acids (the alpha and beta chains) that later become linked together to form the insulin molecule. Figures 2 and 3 are schematics of transposon-based vector constructs containing a proinsulin gene operably-linked to an ovalbumin promoter and ovalbumin protein or an ovomucoid promoter and ovomucoid signal sequence, respectively. In these embodiments, proinsulin is expressed in the oviduct tubular gland cells and then deposited in the egg white. One example of a proinsulin polynucleotide sequence is shown in SEQ ID NO:23, wherein the C-peptide cleavage site spans from Arg at position 31 to Arg at position 65.

[0082] Serum proteins including lipoproteins such as high density lipoprotein (HDL), HDL-Milano and low density lipoprotein, albumin, clotting cascade factors, factor VIII, factor IX, fibrinogen, and globulins are also included in the group of desired proteins of the present invention. Immunoglobulins are one class of desired globulin molecules and include but are not limited to IgG, IgM, IgA, IgD, IgE, IgY, lambda chains, kappa chains and fragments thereof; Fc fragments, and Fab fragments. Desired antibodies include, but are not limited to, naturally occurring antibodies, human antibodies, humanized antibodies, and hybrid antibodies. Genes encoding modified versions of naturally occurring antibodies or fragments thereof and genes encoding artificially designed antibodies or fragments thereof may be incorporated into the transposon-based vectors of the present invention. Desired antibodies also include antibodies with the ability to bind specific ligands, for example, antibodies against proteins associated with cancer-related molecules, such as anti-her 2, or anti-CA125. Accordingly, the present invention encompasses a transposon-based vector containing one or more genes encoding a heavy immunoglobulin (Ig) chain and a light Ig chain. Further, more than one gene encoding for more than one antibody may be administered in one or more transposon-based vectors of the present invention. In this manner, an egg may contain more than one type of antibody in the egg white, the egg yolk or both. In one embodiment, a transposon-based vector contains a heavy Ig chain and a light Ig chain, both operably linked to a promoter.

[0083] Antibodies used as therapeutic reagents include but are not limited to antibodies for use in cancer immunotherapy against specific antigens, or for providing passive immunity to an animal or a human against an infectious disease or a toxic agent. Antibodies used as diagnostic reagents include, but are not limited to antibodies that may be labeled and detected with a detector, for example antibodies with a fluorescent label attached that may be detected following exposure to specific wavelengths. Such labeled antibodies may be primary antibodies directed to a specific antigen, for example, rhodamine-labeled rabbit anti-growth hormone, or may be labeled secondary antibodies, such as fluorescein-labeled goat-anti chicken IgG. Such labeled antibodies are known to one of ordinary skill in the art. Labels useful for attachment to antibodies are also known to one of ordinary skill in the art. Some of these labels are described in the "Handbook of Fluorescent Probes and Research Products", ninth edition, Richard P. Haugland (ed) Molecular Probes, Inc. Eugene, OR), which is incorporated herein in its entirety.

[0084] Antibodies produced with using the present invention may be used as laboratory reagents for numerous applications including radioimmunoassay, western blots, dot blots, ELISA, immunoaffinity columns and other procedures requiring antibodies as known to one of ordinary skill in the art. Such antibodies include primary antibodies, secondary antibodies and tertiary antibodies, which may be labeled or unlabeled.

[0085] Antibodies that may be made with the practice of the present invention include, but are not limited to primary antibodies, secondary antibodies, designer antibodies, anti-protein antibodies, anti-peptide antibodies, anti-DNA antibodies, anti-RNA antibodies, anti-hormone antibodies, anti-hypophysiotropic peptides, antibodies against non-natural antigens, anti-anterior pituitary hormone antibodies, anti-posterior pituitary hormone antibodies, anti-venom antibodies, anti-tumor marker antibodies, antibodies directed against epitopes associated with infectious disease, including, antiviral, anti-bacterial, anti-protozoal, anti-fungal, anti-parasitic, anti-receptor, anti-lipid, anti-phospholipid, anti-growth factor, anti-cytokine, anti-monokine, anti-idiotypic, and anti-accessory (presentation) protein antibodies. Antibodies made with the present invention, as well as light chains or heavy chains, may also be used to inhibit enzyme activity.

[0086] Antibodies that may be produced using the present invention include, but are not limited to, antibodies made against the following proteins: Bovine γ -Globulin, Serum; Bovine IgG, Plasma; Chicken γ -Globulin, Serum; Human γ -Globulin, Serum; Human IgA, Plasma; Human IgA₁, Myeloma; Human IgA₂, Myeloma; Human IgA₂, Plasma; Human IgD, Plasma; Human IgE, Myeloma; Human IgG, Plasma; Human IgG, Fab Fragment, Plasma; Human IgG, F(ab')₂ Fragment, Plasma; Human IgG, Fc Fragment, Plasma; Human IgG₁, Myeloma; Human IgG₂, Myeloma; Human IgG₃, Myeloma; Human IgG₄, Myeloma; Human IgM, Myeloma; Human IgM, Plasma; Human Immunoglobulin, Light Chain κ , Urine; Human Immunoglobulin, Light Chains κ and γ , Plasma; Mouse γ -Globulin, Serum; Mouse IgG, Serum; Mouse IgM, Myeloma; Rabbit γ -Globulin, Serum; Rabbit IgG, Plasma; and Rat γ -Globulin, Serum. In one embodiment, the transposon-based vector comprises the coding sequence of light and heavy chains of a murine monoclonal antibody that shows specificity for human seminoprotein (GenBank Accession numbers AY129006 and AY129304 for the light and heavy chains, respectively).

[0087] A further non-limiting list of antibodies that recognize other antibodies is as follows: Anti-Chicken IgG, heavy (H) & light (L) Chain Specific (Sheep); Anti-Goat γ -Globulin (Donkey); Anti-Goat IgG, Fc Fragment Specific (Rabbit); Anti-Guinea Pig γ -Globulin (Goat); Anti-Human Ig, Light Chain, Type κ Specific; Anti-Human Ig, Light Chain, Type λ

Specific; Anti-Human IgA, α -Chain Specific (Goat); Anti-Human IgA, Fab Fragment Specific; Anti-Human IgA, Fc Fragment Specific; Anti-Human IgA, Secretory; Anti-Human IgE, ϵ -Chain Specific (Goat); Anti-Human IgE, Fc Fragment Specific; Anti-Human IgG, Fc Fragment Specific (Goat); Anti-Human IgG, γ -Chain Specific (Goat); Anti-Human IgG, Fc Fragment Specific; Anti-Human IgG, Fd Fragment Specific; Anti-Human IgG, H & L Chain Specific (Goat); Anti-Human IgG₁, Fc Fragment Specific; Anti-Human IgG₂, Fc Fragment Specific; Anti-Human IgG₂, Fd Fragment Specific; Anti-Human IgG₃, Hinge Specific; Anti-Human IgG₄, Fc Fragment Specific; Anti-Human IgM, Fc Fragment Specific; Anti-Human IgM, μ -Chain Specific; Anti-Mouse IgE, ϵ -Chain Specific; Anti-Mouse γ -Globulin (Goat); Anti-Mouse IgG, γ -Chain Specific (Goat); Anti-Mouse IgG, γ -Chain Specific (Goat) F(ab')₂ Fragment; Anti-Mouse IgG, H & L Chain Specific (Goat); Anti-Mouse IgM, μ -Chain Specific (Goat); Anti-Mouse IgM, H & L Chain Specific (Goat); Anti-Rabbit γ -Globulin (Goat); Anti-Rabbit IgG, Fc Fragment Specific (Goat); Anti-Rabbit IgG, H & L Chain Specific (Goat); Anti-Rat γ -Globulin (Goat); Anti-Rat IgG, H & L Chain Specific; Anti-Rhesus Monkey γ -Globulin (Goat); and, Anti-Sheep IgG, H & L Chain Specific.

[0088] Another non-limiting list of the antibodies that may be produced using the present invention is provided in product catalogs of companies such as Phoenix Pharmaceuticals, Inc. (www.phoenixpeptide.com; 530 Harbor Boulevard, Belmont, CA), Peninsula Labs (San Carlos CA), SIGMA (St Louis, MO www.sigma-aldrich.com), Cappel ICN (Irvine, California, www.icnbiomed.com), and Calbiochem (La Jolla, California, www.calbiochem.com), which are all incorporated herein by reference in their entirety. The polynucleotide sequences encoding these antibodies may be obtained from the scientific literature, from patents, and from databases such as GenBank. Alternatively, one of ordinary skill in the art may design the polynucleotide sequence to be incorporated into the genome by choosing the codons that encode for each amino acid in the desired antibody. Antibodies made by the transgenic animals of the present invention include antibodies that may be used as therapeutic reagents, for example in cancer immunotherapy against specific antigens, as diagnostic reagents and as laboratory reagents for numerous applications including immunoneutralization, radioimmunoassay, western blots, dot blots, ELISA, immunoprecipitation and immunoaffinity columns. Some of these antibodies include, but are not limited to, antibodies which bind the following ligands: adrenomedullin, amylin, calcitonin, amyloid, calcitonin gene-related peptide, cholecystokinin, gastrin, gastric inhibitory peptide, gastrin releasing peptide, interleukin, interferon, cortistatin, somatostatin, endothelin, sarafotoxin, glucagon, glucagon-like peptide, insulin, atrial natriuretic peptide, BNP, CNP, neurokinin, substance P, leptin, neuropeptide Y, melanin concentrating hormone, melanocyte stimulating hormone, orphanin, endorphin, dynorphin, enkephalin, enkephalin, leumorphin, peptide F, PACAP, PACAP-related peptide, parathyroid hormone, urocortin, corticotrophin releasing hormone, PHM, PHI, vasoactive intestinal polypeptide, secretin, ACTH, angiotensin, angiotensin, bombesin, endostatin, bradykinin, FMRF amide, galanin, gonadotropin releasing hormone (GnRH) associated peptide, GnRH, growth hormone releasing hormone, inhibin, granulocyte-macrophage colony stimulating factor (GM-CSF), motilin, neurotensin, oxytocin, vasopressin, osteocalcin, pancreastatin, pancreatic polypeptide, peptide YY, proopiomelanocortin, transforming growth factor, vascular endothelial growth factor, vesicular monoamine transporter, vesicular acetylcholine transporter, ghrelin, NPW, NPB, C3d, prokineticin, thyroid stimulating hormone, luteinizing hormones, follicle stimulating hormone, prolactin, growth hormone, beta-lipotropin, melatonin, kallikriens, kinins, prostaglandins, erythropoietin, p146 (SEQ ID NO:24 amino acid sequence, SEQ ID NO:25, nucleotide sequence), estrogen, testosterone, corticosteroids, mineralocorticoids, thyroid hormone, thymic hormones, connective tissue proteins, nuclear proteins, actin, avidin, activin, agrin, albumin, and prohormones, propeptides, splice variants, fragments and analogs thereof.

[0089] The following is yet another non-limiting list of antibodies that can be produced by the methods of present invention: abciximab (ReoPro), abciximab antiplatelet aggregation monoclonal antibody, anti-CD11a (hu1124), anti-CD18 antibody, anti-CD20 antibody, anti-cytomegalovirus (CMV) antibody, anti-digoxin antibody, anti-hepatitis B antibody, anti-HER-2 antibody, anti-idiotypic antibody to GD3 glycolipid, anti-IgE antibody, anti-IL-2R antibody, antimetastatic cancer antibody (mAb 17-1A), anti-rabies antibody, anti-respiratory syncytial virus (RSV) antibody, anti-Rh antibody, anti-TCR, anti-TNF antibody, anti-VEGF antibody and fab fragment thereof, rattlesnake venom antibody, black widow spider venom antibody, coral snake venom antibody, antibody against very late antigen-4 (VLA-4), C225 humanized antibody to EGF receptor, chimeric (human & mouse) antibody against TNF α , antibody directed against GPIIb/IIIa receptor on human platelets, gamma globulin, anti-hepatitis B immunoglobulin, human anti-D immunoglobulin, human antibodies against *S aureus*, human tetanus immunoglobulin, humanized antibody against the epidermal growth receptor-2, humanized antibody against the α subunit of the interleukin-2 receptor, humanized antibody CTLA4IG, humanized antibody to the IL-2 R α -chain, humanized anti-CD40-ligand monoclonal antibody (5c8), humanized mAb against the epidermal growth receptor-2, humanized mAb to rous sarcoma virus, humanized recombinant antibody (IgG1k) against respiratory syncytial virus (RSV), lymphocyte immunoglobulin (anti-thymocyte antibody), lymphocyte immunoglobulin, mAb against factor VII, MDX-210 bi-specific antibody against HER-2, MDX-22, MDX-220 bi-specific antibody against TAG-72 on tumors, MDX-33 antibody to Fc γ R1 receptor, MDX-447 bi-specific antibody against EGF receptor, MDX-447 bispecific humanized antibody to EGF receptor, MDX-RA immunotoxin (ricin A linked) antibody, Medi-507 antibody (humanized form of BTI-322) against CD2 receptor on T-cells, monoclonal antibody LDP-02, muromonab-CD3(OKT3) antibody, OKT3 ("muromonab-CD3") antibody, PRO 542 antibody, ReoPro ("abciximab") antibody, and TNP-IgG fusion protein.

[0090] The antibodies prepared using the methods of the present invention may also be designed to possess specific labels that may be detected through means known to one of ordinary skill in the art. The antibodies may also be designed to possess specific sequences useful for purification through means known to one of ordinary skill in the art. Specialty antibodies designed for binding specific antigens may also be made in transgenic animals using the transposon-based vectors of the present invention.

[0091] production of a monoclonal antibody using the transposon-based vectors of the present invention can be accomplished in a variety of ways. In one embodiment, two vectors may be constructed: one that encodes the light chain, and a second vector that encodes the heavy chain of the monoclonal antibody. These vectors may then be incorporated into the genome of the target animal by methods disclosed herein. In an alternative embodiment, the sequences encoding light and heavy chains of a monoclonal antibody may be included on a single DNA construct. For example, the coding sequence of light and heavy chains of a murine monoclonal antibody that show specificity for human seminoprotein can be expressed using transposon-based constructs of the present invention (GenBank Accession numbers AY129006 and AY129304 for the light and heavy chains, respectively).

[0092] Further included in the present invention are proteins and peptides synthesized by the immune system including those synthesized by the thymus, lymph nodes, spleen, and the gastrointestinal associated lymph tissues (GALT) system. The immune system proteins and peptides proteins that can be made in transgenic animals using the transposon-based vectors of the present invention include, but are not limited to, alpha-interferon, beta-interferon, gamma-interferon, alpha-interferon A, alpha-interferon 1, G-CSF, GM-CSF, interleukin-1 (IL-1), IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, TNF- α , and TNF- β . Other cytokines included in the present invention include cardiotrophin, stromal cell derived factor, macrophage derived chemokine (MDC), melanoma growth stimulatory activity (MGSA), macrophage inflammatory proteins 1 alpha (MIP-1 alpha), 2, 3 alpha, 3 beta, 4 and 5.

[0093] Lytic peptides such as p146 are also included in the desired molecules of the present invention. In one embodiment, the p146 peptide comprises an amino acid sequence of SEQ ID NO:24. The present invention also encompasses a transposon-based vector comprising a p146 nucleic acid comprising a polynucleotide sequence of SEQ ID NO:25.

[0094] Enzymes are another class of proteins that may be made through the use of the transposon-based vectors of the present invention. Such enzymes include but are not limited to adenosine deaminase, alpha-galactosidase, cellulase, collagenase, dnasel, hyaluronidase, lactase, L-asparaginase, pancreatin, papain, streptokinase B, subtilisin, superoxide dismutase, thrombin, trypsin, urokinase, fibrinolysin, glucocerebrosidase and plasminogen activator. In some embodiments wherein the enzyme could have deleterious effects, additional amino acids and a protease cleavage site are added to the carboxy end of the enzyme of interest in order to prevent expression of a functional enzyme. Subsequent digestion of the enzyme with a protease results in activation of the enzyme.

[0095] Extracellular matrix proteins are one class of desired proteins that may be made through the use of the present invention. Examples include but are not limited to collagen, fibrin, elastin, laminin, and fibronectin and subtypes thereof. Intracellular proteins and structural proteins are other classes of desired proteins in the present invention.

[0096] Growth factors are another desired class of proteins that may be made through the use of the present invention and include, but are not limited to, transforming growth factor- α ("TGF- α "), transforming growth factor- β (TGF- β), platelet-derived growth factors (PDGF), fibroblast growth factors (FGF), including FGF acidic isoforms 1 and 2, FGF basic form 2 and FGF 4, 8, 9 and 10, nerve growth factors (NGF) including NGF 2.5s, NGF 7.0s and beta NGF and neurotrophins, brain derived neurotrophic factor, cartilage derived factor, growth factors for stimulation of the production of red blood cells, growth factors for stimulation of the production of white blood cells, bone growth factors (BGF), basic fibroblast growth factor, vascular endothelial growth factor (VEGF), granulocyte colony stimulating factor (G-CSF), insulin like growth factor (IGF) I and II, hepatocyte growth factor, glial neurotrophic growth factor (GDNF), stem cell factor (SCF), keratinocyte growth factor (KGF), transforming growth factors (TGF), including TGFs alpha, beta, beta1, beta2, beta3, skeletal growth factor, bone matrix derived growth factors, bone derived growth factors, erythropoietin (EPO) and mixtures thereof.

[0097] Another desired class of proteins that may be made may be made through the use of the present invention include, but are not limited to, leptin, leukemia inhibitory factor (LIF), tumor necrosis factor alpha and beta, ENBREL, angiostatin, endostatin, thrombospondin, osteogenic protein-1, bone morphogenetic proteins 2 and 7, osteonectin, somatomedin-like peptide, and osteocalcin.

[0098] Yet another desired class of proteins are blood proteins or clotting cascade protein including albumin, Prekallikrein, High molecular weight kininogen (HMWK) (contact activation cofactor, Fitzgerald, Flaujeac Williams factor), Factor I (Fibrinogen), Factor II (prothrombin), Factor III (Tissue Factor), Factor IV (calcium), Factor V (proaccelerin, labile factor, accelerator (Ac-) globulin), Factor VI (Va) (accelerin), Factor VII (proconvertin), serum prothrombin conversion accelerator (SPCA), cothromboplastin), Factor VIII (antihemophilic factor A, antihemophilic globulin (AHG)), Factor IX (Christmas Factor, antihemophilic factor B, plasma thromboplastin component (PTC)), Factor X (Stuart-Prower Factor), Factor XI (Plasma thromboplastin antecedent (PTA)), Factor XII (Hageman Factor), Factor XIII (rotransglutaminase, fibrin stabilizing factor (FSF), fibrinoligase), von Willebrand factor, Protein C, Protein S, Thrombomodulin, Antithrombin III.

[0099] A non-limiting list of the peptides and proteins that may be made may be made through the use of the present invention is provided in product catalogs of companies such as Phoenix Pharmaceuticals, Inc. (www.phoenixpeptide.com; 530 Harbor Boulevard, Belmont, CA), Peninsula Labs (San Carlos CA), SIGMA, (St. Louis, MO www.sigma-aldrich.com), Cappel ICN (Irvine, California, www.icnbiomed.com), and Calbiochem (La Jolla, California, www.calbiochem.com). The polynucleotide sequences encoding these proteins and peptides of interest may be obtained from the scientific literature, from patents, and from databases such as GenBank. Alternatively, one of ordinary skill in the art may design the polynucleotide sequence to be incorporated into the genome by choosing the codons that encode for each amino acid in the desired protein or peptide.

[0100] Some of these desired proteins or peptides that may be made through the use of the present invention include but are not limited to the following: adrenomedullin, amylin, calcitonin, amyloid, calcitonin gene-related peptide, cholecystokinin, gastrin, gastric inhibitory peptide, gastrin releasing peptide, interleukin, interferon, cortistatin, somatostatin, endothelin, sarafotoxin, glucagon, glucagon-like peptide, insulin, atrial natriuretic peptide, BNP, CNP, neurokinin, substance P, leptin, neuropeptide Y, melanin concentrating hormone, melanocyte stimulating hormone, orphanin, endorphin, dynorphin, enkephalin, leumorphin, peptide F, PACAP, PACAP-related peptide, parathyroid hormone, urocortin, corticotrophin releasing hormone, PHM, PHI, vasoactive intestinal polypeptide, secretin, ACTH, angiotensin, angiotensin, bombesin, endostatin, bradykinin, FMRF amide, galanin, gonadotropin releasing hormone (GnRH) associated peptide, GnRH, growth hormone releasing hormone, inhibin, granulocyte-macrophage colony stimulating factor (GM-CSF), motilin, neurotensin, oxytocin, vasopressin, osteocalcin, pancreastatin, pancreatic polypeptide, peptide YY, proopiomelanocortin, transforming growth factor, vascular endothelial growth factor, vesicular monoamine transporter, vesicular acetylcholine transporter, ghrelin, NPW, NPB, C3d, prokineticin, thyroid stimulating hormones, luteinizing hormone, follicle stimulating hormone, prolactin, growth hormone, beta-lipotropin, melatonin, kallikriens, kinins, prostaglandins, erythropoietin, p146 (SEQ ID NO:24, amino acid sequence, SEQ ID NO:25, nucleotide sequence), thymic hormones, connective tissue proteins, nuclear proteins, actin, avidin, activin, agrin, albumin, apolipoproteins, apolipoprotein A, apolipoprotein B, and prohormones, propeptides, splice variants, fragments and analogs thereof.

[0101] Other desired proteins that may be made by the method of the present invention include bacitracin, polymyxin b, vancomycin, cyclosporine, anti-RSV antibody, alpha-1 antitrypsin (AAT), anti-cytomegalovirus antibody, anti-hepatitis antibody, anti-inhibitor coagulant complex, anti-rabies antibody, anti-Rh(D) antibody, adenosine deaminase, anti-digoxin antibody, antivenin crotalidae (rattlesnake venom antibody), antivenin latrodectus (black widow spider venom antibody), antivenin micrurus (coral snake venom antibody), aprotinin, corticotropin (ACTH), diphtheria antitoxin, lymphocyte immune globulin (anti-thymocyte antibody), protamine, thyrotropin, capreomycin; α -galactosidase, gramicidin, streptokinase, tetanus toxoid, tyrothricin, IGF-1, proteins of varicella vaccine, anti-TNF antibody, anti-IL-2r antibody, anti-HER-2 antibody, OKT3 ("muromonab-CD3") antibody, TNF-IgG fusion protein, ReoPro ("abciximab") antibody, ACTH fragment 1-24, desmopressin, gonadotropin-releasing hormone, histrelin, leuprolide, lypressin, nafarelin, peptide that binds GPIIb/GPIIIa on platelets (integrilin), goserelin, capreomycin, colistin, anti-respiratory syncytial virus, lymphocyte immune globulin (Thymoglovin, Atgam), panorex, alpha-antitrypsin, botulinin, lung surfactant protein, tumor necrosis receptor-IgG fusion protein (enbrel), gonadorelin, proteins of influenza vaccine, proteins of rotavirus vaccine, proteins of haemophilus b conjugate vaccine, proteins of poliovirus vaccine, proteins of pneumococcal conjugate vaccine, proteins of meningococcal C vaccine, proteins of influenza vaccine, megakaryocyte growth and development factor (MGDF), neuroimmunophilin ligand-A (NIL-A), brain-derived neurotrophic factor (BDNF), glial cell line-derived neurotrophic factor (GDNF), leptin (native), leptin B, leptin C, IL-1RA (interleukin-1RA), R-568, novel erythropoiesis-stimulating protein (NESP), humanized mAb to rous sarcoma virus (MEDI-493), glutamyl-tryptophan dipeptide IM862, LFA-3TIP immunosuppressive, humanized anti-CD40-ligand monoclonal antibody (5c8), gelsolin enzyme, tissue factor pathway inhibitor (TFPI), proteins of meningitis B vaccine, antimetastatic cancer antibody (mAb 17-1A), chimeric (human & mouse) mAb against TNF α , mAb against factor VII, relaxin, capreomycin, glycopeptide (LY333328), recombinant human activated protein C (rhAPC), humanized mAb against the epidermal growth receptor-2, altepase, anti-CD20 antigen, C2B8 antibody, insulin-like growth factor-1, atrial natriuretic peptide (anaritide), tenectaplastase, anti-CD11a antibody (hu 1124), anti-CD18 antibody, mAb LDP-02, anti-VEGF antibody, fab fragment of anti-VEGF Ab, AP02 ligand (tumor necrosis factor-related apoptosis-inducing ligand), rTGF- β (transforming growth factor- β), alpha-antitrypsin, ananain (a pineapple enzyme), humanized mAb CTLA4IG, PRO 542 (mAb), D2E7 (mAb), calf intestine alkaline phosphatase, α -L-iduronidase, α -L-galactosidase (humanglutamic acid decarboxylase, acid sphingomyelinase, bone morphogenetic protein-2 (rhBMP-2), proteins of HIV vaccine, T cell receptor (TCR) peptide vaccine, TCR peptides, V beta 3 and V beta 13.1. (IR502), (In501), BI 1050/1272 mAb against very late antigen-4 (VLA-4), C225 humanized mAb to EGF receptor, anti-idiotypic antibody to GD3 glycolipid, antibacterial peptide against *H. pylori*, MDX-447 bispecific humanized mAb to EGF receptor, anti-cytomegalovirus (CMV), Medi-491 B19 parvovirus vaccine, humanized recombinant mAb (IgG1k) against respiratory syncytial virus (RSV), urinary tract infection vaccine (against "pili" on *Escherichia coli* strains), proteins of lyme disease vaccine against *B. burgdorferi* protein (DbpA), proteins of Medi-501 human papilloma virus-11 vaccine (HPV), *Streptococcus pneumoniae* vaccine, Medi-507 mAb (humanized form of BTI-322) against CD2 receptor on T-cells, MDX-33 mAb to Fc γ R1 receptor, MDX-RA immunotoxin (ricin A linked) mAb, MDX-210 bi-specific mAb against HER-2, MDX-

447 bi-specific mAb against EGF receptor, MDX-22, MDX-220 bi-specific mAb against TAG-72 on tumors, colony-stimulating factor (CSF) (molgramostim), humanized mAb to the IL-2 R α -chain (basiliximab), mAb to IgE (IGE 025A), myelin basic protein-altered peptide (MSP771A), humanized mAb against the epidermal growth receptor-2, humanized mAb against the α subunit of the interleukin-2 receptor, low molecular weight heparin, anti-hemophilic factor, and bactericidal/permeability-increasing protein (r-BPI).

[0102] The peptides and proteins made using the present invention may be labeled using labels and techniques known to one of ordinary skill in the art. Some of these labels are described in the "Handbook of Fluorescent Probes and Research Products", ninth edition, Richard P. Haugland (ed) Molecular Probes, Inc. Eugene, OR), which is incorporated herein in its entirety. Some of these labels may be genetically engineered into the polynucleotide sequence for the expression of the selected protein or peptide. The peptides and proteins may also have label-incorporation "handles" incorporated to allow labeling of an otherwise difficult or impossible to label protein.

[0103] It is to be understood that the various classes of desired peptides and proteins, as well as specific peptides and proteins described in this section may be modified as described below by inserting selected codons for desired amino acid substitutions into the gene incorporated into the transgenic animal.

[0104] Also, reference is made to the production of molecules other than proteins and peptides including, but not limited to, lipoproteins such as high density lipoprotein (HDL), HDL-Milano, and low density lipoprotein, lipids, carbohydrates, siRNA and ribozymes. In such cases, a gene of interest encodes a nucleic acid molecule or a protein that directs production of the desired molecule.

[0105] Further, reference is made to the use of inhibitory molecules to inhibit endogenous (i.e., non-vector) protein production. These inhibitory molecules include antisense nucleic acids, siRNA and inhibitory proteins.

The endogenous protein whose expression is inhibited may be an egg white protein including, but not limited to ovalbumin, ovotransferrin, and ovomucin.

[0106] A transposon-based vector containing an ovalbumin DNA sequence, that upon transcription forms a double stranded RNA molecule, may be transfected into an animal such as a bird and the bird's production of endogenous ovalbumin protein is reduced by the interference RNA mechanism (RNAi).

A transposon-based vector may encode an inhibitory RNA molecule that inhibits the expression of more than one egg white protein. One exemplary construct is provided in Figure 4 wherein "Ovgen" indicates approximately 60 base pairs of an ovalbumin gene, "Ovotrans" indicates approximately 60 base pairs of an ovotransferrin gene and "Ovomucin" indicates approximately 60 base pairs of an ovomucin gene. These ovalbumin, ovotransferrin and ovomucin can be from any avian species, and in some cases, are from a chicken or quail. The term "pro" indicates the pro portion of a prepro sequence. One exemplary prepro sequence is that of cecropin and comprising base pairs 563-733 of the Cecropin cap site and Prepro provided in Genbank accession number X07404. Additional cecropin prepro and pro sequences are provided in SEQ ID NO:48, SEQ ID NO:49, SEQ ID NO:50, and SEQ ID NO:51. Additionally, inducible knockouts or knockdowns of the endogenous protein may be created to achieve a reduction or inhibition of endogenous protein production. Endogenous egg white production can be inhibited in an avian at any time, but is preferably inhibited preceding, or immediately preceding, the harvest of eggs.

Modified Desired Proteins and Peptides

[0107] "Proteins," "peptides," "polypeptides" and "oligopeptides" are chains of amino acids (typically L-amino acids) whose alpha carbons are linked through peptide bonds formed by a condensation reaction between the carboxyl group of the alpha carbon of one amino acid and the amino group of the alpha carbon of another amino acid. The terminal amino acid at one end of the chain (i.e., the amino terminal) has a free amino group, while the terminal amino acid at the other end of the chain (i.e., the carboxy terminal) has a free carboxyl group. As such, the term "amino terminus" (abbreviated N-terminus) refers to the free alpha-amino group on the amino acid at the amino terminal of the protein, or to the alpha-amino group (imino group when participating in a peptide bond) of an amino acid at any other location within the protein. Similarly, the term "carboxy terminus" (abbreviated C-terminus) refers to the free carboxyl group on the amino acid at the carboxy terminus of a protein, or to the carboxyl group of an amino acid at any other location within the protein.

[0108] Typically, the amino acids making up a protein are numbered in order, starting at the amino terminal and increasing in the direction toward the carboxy terminal of the protein. Thus, when one amino acid is said to "follow" another, that amino acid is positioned closer to the carboxy terminal of the protein than the preceding amino acid.

[0109] The term "residue" is used herein to refer to an amino acid (D or L) or an amino acid mimetic that is incorporated into a protein by an amide bond. As such, the amino acid may be a naturally occurring amino acid or, unless otherwise limited, may encompass known analogs of natural amino acids that function in a manner similar to the naturally occurring amino acids (i.e., amino acid mimetics). Moreover, an amide bond mimetic includes peptide backbone modifications well known to those skilled in the art.

[0110] Furthermore, one of skill will recognize that, as mentioned above, individual substitutions, deletions or additions

which alter, add or delete a single amino acid or a small percentage of amino acids (typically less than about 5%, more typically less than about 1%) in an encoded sequence are conservatively modified variations where the alterations result in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. The following six groups each contain amino acids that are

5

- 1) Alanine (A), Serine (S), Threonine (T);
- 2) Aspartic acid (D), Glutamic acid (B);
- 3) Asparagine (N), Glutamine (Q);
- 4) Arginine (R), Lysine (K);
- 5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and
- 6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

10

[0111] A conservative substitution is a substitution in which the substituting amino acid (naturally occurring or modified) is structurally related to the amino acid being substituted, i.e., has about the same size and electronic properties as the amino acid being substituted. Thus, the substituting amino acid would have the same or a similar functional group in the side chain as the original amino acid. A "conservative substitution" also refers to utilizing a substituting amino acid which is identical to the amino acid being substituted except that a functional group in the side chain is protected with a suitable protecting group.

15

[0112] Suitable protecting groups are described in Green and Wuts, "Protecting Groups in Organic Synthesis", John Wiley and Sons, Chapters 5 and 7, 1991, the teachings of which are incorporated herein by reference. Preferred protecting groups are those which facilitate transport of the peptide through membranes, for example, by reducing the hydrophilicity and increasing the lipophilicity of the peptide, and which can be cleaved, either by hydrolysis or enzymatically (Ditter et al., 1968. J. Pharm. Sci. 57:783; Ditter et al., 1968. J. Pharm. Sci. 57:828; Ditter et al., 1969. J. Pharm. Sci. 58:557; King et al., 1987. Biochemistry 26:2294; Lindberg et al., 1989. Drug Metabolism and Disposition 17:311; Tunek et al., 1988. Biochem. Pharm. 37:3867; Anderson et al., 1985 Arch. Biochem. Biophys. 239:538; and Singhal et al., 1987. FASEB J. 1:220). Suitable hydroxyl protecting groups include ester, carbonate and carbamate protecting groups. Suitable amine protecting groups include acyl groups and alkoxy or aryloxy carbonyl groups, as described above for N-terminal protecting groups. Suitable carboxylic acid protecting groups include aliphatic, benzyl and aryl esters, as described below for C-terminal protecting groups. In one embodiment, the carboxylic acid group in the side chain of one or more glutamic acid or aspartic acid residues in a peptide of the present invention is protected, preferably as a methyl, ethyl, benzyl or substituted benzyl ester, more preferably as a benzyl ester.

20

25

30

[0113] Provided below are groups of naturally occurring and modified amino acids in which each amino acid in a group has similar electronic and steric properties. Thus, a conservative substitution can be made by substituting an amino acid with another amino acid from the same group. It is to be understood that these groups are non-limiting, i.e. that there are additional modified amino acids which could be included in each group.

35

- | | |
|-----------|--|
| Group I | includes leucine, isoleucine, valine, methionine and modified amino acids having the following side chains: ethyl, n-propyl n-butyl. Preferably, Group I includes leucine, isoleucine, valine and methionine. |
| Group II | includes glycine, alanine, valine and a modified amino acid having an ethyl side chain. Preferably, Group II includes glycine and alanine. |
| Group III | includes phenylalanine, phenylglycine, tyrosine, tryptophan, cyclohexylmethyl glycine, and modified amino residues having substituted benzyl or phenyl side chains. Preferred substituents include one or more of the following: halogen, methyl, ethyl, nitro, -NH ₂ , methoxy, methoxy and -CN. Preferably, Group III includes phenylalanine, tyrosine and tryptophan. |
| Group IV | includes glutamic acid, aspartic acid, a substituted or unsubstituted aliphatic, aromatic or benzylic ester of glutamic or aspartic acid (e.g., methyl, ethyl, n-propyl iso-propyl, cyclohexyl, benzyl or substituted benzyl), glutamine, asparagine, -CO-NH- alkylated glutamine or asparagines (e.g., methyl, ethyl, n-propyl and iso-propyl) and modified amino acids having the side chain -(CH ₂) ₃ -COOH, an ester thereof (substituted or unsubstituted aliphatic, aromatic or benzylic ester), an amide thereof and a substituted or unsubstituted N-alkylated amide thereof. Preferably, Group IV includes glutamic acid, aspartic acid, methyl aspartate, ethyl aspartate, benzyl aspartate and methyl glutamate, ethyl glutamate and benzyl glutamate, glutamine and asparagine. |
| Group V | includes histidine, lysine, ornithine, arginine, N-nitroarginine, β-cycloarginine, γ-hydroxyarginine, N-amidinocitrulline and 2-amino-4-guanidinobutanoic acid, homologs of lysine, homologs of arginine and homologs of ornithine. Preferably, Group V includes histidine, lysine, arginine and ornithine. A homolog of an amino acid includes from 1 to about 3 additional or subtracted methylene units in the side chain. |
| Group VI | includes serine, threonine, cysteine and modified amino acids having C1-C5 straight or branched alkyl side |

40

45

50

55

chains substituted with -OH or -SH, for example, $-\text{CH}_2\text{CH}_2\text{OH}$, $-\text{CH}_2\text{CH}_2\text{CH}_2\text{OH}$ or $-\text{CH}_2\text{CH}_2\text{OHCH}_3$. Preferably, Group VI includes serine, cysteine or threonine.

[0114] In another aspect, suitable substitutions for amino acid residues include "severe" substitutions. A "severe substitution" is a substitution in which the substituting amino acid (naturally occurring or modified) has significantly different size and/or electronic properties compared with the amino acid being substituted. Thus, the side chain of the substituting amino acid can be significantly larger (or smaller) than the side chain of the amino acid being substituted and/or can have functional groups with significantly different electronic properties than the amino acid being substituted. Examples of severe substitutions of this type include the substitution of phenylalanine or cyclohexylmethyl glycine for alanine, isoleucine for glycine, a D amino acid for the corresponding L amino acid, or $-\text{NH}-\text{CH}[(\text{-CH}_2)_5-\text{COOH}]-\text{CO}-$ for aspartic acid. Alternatively, a functional group may be added to the side chain, deleted from the side chain or exchanged with another functional group. Examples of severe substitutions of this type include adding of valine, leucine or isoleucine, exchanging the carboxylic acid in the side chain of aspartic acid or glutamic acid with an amine, or deleting the amine group in the side chain of lysine or ornithine. In yet another alternative, the side chain of the substituting amino acid can have significantly different steric and electronic properties than the functional group of the amino acid being substituted. Examples of such modifications include tryptophan for glycine, lysine for aspartic acid and $-(\text{CH}_2)_4\text{COOH}$ for the side chain of serine. These examples are not meant to be limiting.

[0115] In another embodiment, for example in the synthesis of a peptide 26 amino acids in length, the individual amino acids may be substituted according in the following manner.

AA₁ is serine, glycine, alanine, cysteine or threonine;
 AA₂ is alanine, threonine, glycine, cysteine or serine;
 AA₃ is valine, arginine, leucine, isoleucine, methionine, ornithine, lysine, N-nitroarginine, β -cycloarginine, γ -hydroxyarginine, N-amidinocitruline or 2-amino-4-guanidinobutanoic acid;
 AA₄ is proline, leucine, valine, isoleucine or methionine;
 AA₅ is tryptophan, alanine, phenylalanine, tyrosine or glycine;
 AA₆ is serine, glycine, alanine, cysteine or threonine;
 AA₇ is proline, leucine, valine, isoleucine or methionine;
 AA₈ is alanine, threonine, glycine, cysteine or serine;
 AA₉ is alanine, threonine, glycine, cysteine or serine;
 AA₁₀ is leucine, isoleucine, methionine or valine;
 AA₁₁ is serine, glycine, alanine, cysteine or threonine;
 AA₁₂ is leucine, isoleucine, methionine or valine;
 AA₁₃ is leucine, isoleucine, methionine or valine;
 AA₁₄ is glutamine, glutamic acid, aspartic acid, asparagine, or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;
 AA₁₅ is arginine, N-nitroarginine, β -cycloarginine, γ -hydroxy-arginine, N-amidinocitruline or 2-amino-4-guanidinobutanoic acid;
 AA₁₆ is proline, leucine, valine, isoleucine or methionine;
 AA₁₇ is serine, glycine, alanine, cysteine or threonine;
 AA₁₈ is glutamic acid, aspartic acid, asparagine, glutamine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;
 AA₁₉ is aspartic acid, asparagine, glutamic acid, glutamine, leucine, valine, isoleucine, methionine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;
 AA₂₀ is valine, arginine, leucine, isoleucine, methionine, ornithine, lysine, N-nitroarginine, β -cycloarginine, γ -hydroxyarginine, N-amidinocitruline or 2-amino-4-guanidinobutanoic acid;
 AA₂₁ is alanine, threonine, glycine, cysteine or serine;
 AA₂₂ is alanine, threonine, glycine, cysteine or serine;
 AA₂₃ is histidine, serine, threonine, cysteine, lysine or ornithine;
 AA₂₄ is threonine, aspartic acid, serine, glutamic acid or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid;
 AA₂₅ is asparagine, aspartic acid, glutamic acid, glutamine, leucine, valine, isoleucine, methionine or a substituted or unsubstituted aliphatic or aryl ester of glutamic acid or aspartic acid; and
 AA₂₆ is cysteine, histidine, serine, threonine, lysine or ornithine.

[0116] It is to be understood that these amino acid substitutions may be made for longer or shorter peptides than the 26 mer in the preceding example above, and for proteins.

[0117] In one embodiment of the present invention, codons for the first several N-terminal amino acids of the trans-

posase are modified such that the third base of each codon is changed to an A or a T without changing the corresponding amino acid. It is preferable that between approximately 1 and 20, more preferably 3 and 15, and most preferably between 4 and 12 of the first N-terminal codons of the gene of interest are modified such that the third base of each codon is changed to an A or a T without changing the corresponding amino acid. In one embodiment, the first ten N-terminal codons of the gene of interest are modified in this manner.

[0118] When several desired proteins, protein fragments or peptides are encoded in the gene of interest to be incorporated into the genome, one of skill in the art will appreciate that the proteins, protein fragments or peptides may be separated by a spacer molecule such as, for example, a peptide, consisting of one or more amino acids. Generally, the spacer will have no specific biological activity other than to join the desired proteins, protein fragments or peptides together, or to preserve some minimum distance or other spatial relationship between them. However, the constituent amino acids of the spacer may be selected to influence some property of the molecule such as the folding, net charge, or hydrophobicity. The spacer may also be contained within a nucleotide sequence with a purification handle or be flanked by cleavage sites, such as proteolytic cleavage sites.

[0119] Such polypeptide spacers may have from about 5 to about 40 amino acid residues. The spacers in a polypeptide are independently chosen, but are preferably all the same. The spacers should allow for flexibility of movement in space and are therefore typically rich in small amino acids, for example, glycine, serine, proline or alanine. Preferably, peptide spacers contain at least 60%, more preferably at least 80% glycine or alanine. In addition, peptide spacers generally have little or no biological and antigenic activity. Preferred spacers are $(\text{Gly-Pro-Gly-Gly})_x$ (SEQ ID NO:26) and $(\text{Gly}_4\text{-Ser})_y$, wherein x is an integer from about 3 to about 9 and y is an integer from about 1 to about 8. Specific examples of suitable spacers include

$(\text{Gly-Pro-Gly-Gly})_3$

SEQ ID NO:27 Gly Pro Gly Gly Gly Pro Gly Gly Gly Pro Gly Gly

$(\text{Gly}_4\text{-Ser})_3$

SEQ ID NO:28 Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser

or $(\text{Gly}_4\text{-Ser})_4$

**SEQ ID NO:29 Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser
Gly Gly Gly Gly Ser.**

[0120] Nucleotide sequences encoding for the production of residues which may be useful in purification of the expressed recombinant protein may also be built into the vector. Such sequences are known in the art and include then glutathione binding domain from glutathione S-transferase, polylysine, hexa-histidine or other cationic amino acids, thioredoxin, hemagglutinin antigen and maltose binding protein.

[0121] Additionally, nucleotide sequences may be inserted into the gene of interest to be incorporated so that the protein or peptide can also include from one to about six amino acids that create signals for proteolytic cleavage. In this manner, if a gene is designed to make one or more peptides or proteins of interest in the transgenic animal, specific nucleotide sequences encoding for amino acids recognized by enzymes may be incorporated into the gene to facilitate cleavage of the large protein or peptide sequence into desired peptides or proteins or both. For example, nucleotides encoding a proteolytic cleavage site can be introduced into the gene of interest so that a signal sequence can be cleaved from a protein or peptide encoded by the gene of interest. Nucleotide sequences encoding other amino acid sequences which display pH sensitivity or chemical sensitivity may also be added to the vector to facilitate separation of the signal sequence from the peptide or protein of interest.

[0122] Proteolytic cleavage sites include cleavage sites recognized by exopeptidases such as carboxypeptidase A, carboxypeptidase B, aminopeptidase I, and dipeptidylaminopeptidase; endopeptidases such as trypsin, V8-protease, enterokinase, factor Xa, collagenase, endoproteinase, subtilisin, and thrombin; and proteases such as Protease 3C IgA protease (Igase) Rhinovirus 3C(preScission)protease. Chemical cleavage sites are also included in the definition of cleavage site as used herein. Chemical cleavage sites include, but are not limited to, site cleaved by cyanogen bromide, hydroxylamine, formic acid, and acetic acid.

[0123] In one embodiment of the present invention, a TAG sequence is linked to the gene of interest. The TAG sequence serves three purposes: 1) it allows free rotation of the peptide or protein to be isolated so there is no interference from the native protein or signal sequence, i.e. vitellogenin, 2) it provides a "purification handle" to isolate the protein using column purification, and 3) it includes a cleavage site to remove the desired protein from the signal and purification sequences. Accordingly, as used herein, a TAG sequence includes a spacer sequence, a purification handle and a cleavage site. The spacer sequences in the TAG proteins contain one or more repeats shown in SEQ ID NO:30. A preferred spacer sequence comprises the sequence provided in SEQ ID NO:31. One example of a purification handle is the gp41 hairpin loop from HIV I. Exemplary gp41 polynucleotide and polypeptide sequences are provided in SEQ ID

NO:32 and SEQ ID NO:33, respectively. However, it should be understood that any antigenic region may be used as a purification handle, including any antigenic region of gp41. Preferred purification handles are those that elicit highly specific antibodies. Additionally, the cleavage site can be any protein cleavage site known to one of ordinary skill in the art and includes an enterokinase cleavage site comprising the Asp Asp Asp Asp Lys sequence (SEQ ID NO:34) and a furin cleavage site. Constructs containing a TAG sequence are shown in Figures 2 and 3. In one embodiment of the present invention, the TAG sequence comprises a polynucleotide sequence of SEQ ID NO:35.

Methods of Administering Transposon-Based Vectors

[0124] The present invention includes methods of administering the transposon-based vectors to a bird. The present invention makes also reference to methods of producing a transgenic animal wherein a gene of interest is incorporated into the germline of the animal and methods of producing a transgenic animal wherein a gene of interest is incorporated into cells other than the germline cells (somatic cells) of the animal. The transposon-based vectors of the present invention are administered to an oviduct or an ovary via any method known to those of skill in the art. According to present claim 1 reproductive organ means an oviduct or an ovary.

[0125] In some embodiments, a transposon-based vector is directly administered to the oviduct or ovary. Direct administration encompasses injection into the organ, and in a preferred embodiment; a transposon-based vector is injected into the lumen of the oviduct, and more preferably, the lumen of the magnum or the infundibulum of the oviduct. The transposon-based vectors may additionally or alternatively be placed in an artery supplying the reproductive organ. Administering the vectors to the artery supplying the ovary results in transfection of follicles and oocytes in the ovary to create a germline transgenic animal. Alternatively, supplying the vectors through an artery leading to the oviduct would preferably transfect the tubular gland and epithelial cells. Such transfected cells could manufacture a desired protein or peptide for deposition in the egg white. In one embodiment, a transposon-based vector is administered into the lumen of the magnum or the infundibulum of the oviduct and to an artery supplying the oviduct. Indirect administration to the oviduct epithelium may occur through the cloaca. Direct administration into the mammary gland comprises introduction into the duct system of the mammary gland.

[0126] Administration of transposon-based vectors may occur in arteries supplying the ovary and or through direct intrathecal administration into the ovary through injection.

[0127] The transposon-based vectors may be administered in a single administration, multiple administrations, continuously, or intermittently. The transposon-based vectors may be administered by injection, via a catheter, an osmotic mini-pump or any other method. In some embodiments, the transposon-based vector is administered to an animal in multiple administrations, each administration containing the vector and a different transfecting reagent.

[0128] The transposon-based vectors may be administered to the bird at any point during the lifetime of the bird however, it is preferable that the vectors are administered prior to the bird reaching sexual maturity. The transposon-based vectors are preferably administered to a chicken between approximately 14 and 16 weeks of age and to a quail between approximately 5 and 10 weeks of age, more preferably 5 and 8 weeks of age, and most preferably between 5 and 6 weeks of age, when standard poultry rearing practices are used. The vectors may be administered at earlier ages when exogenous hormones are used to induce early sexual maturation in the bird. In some embodiments, the transposon-based vector is administered to a bird following an increase in proliferation of the oviduct epithelial cells and/or the tubular gland cells. Such an increase in proliferation normally follows an influx of reproductive hormones in the area of the oviduct. When the bird is an avian, the transposon-based vector is administered following an increase in proliferation of the oviduct epithelial cells and before the avian begins to produce egg white constituents.

[0129] In a preferred embodiment, the bird is an avian. In one embodiment, between approximately 1 and 150 μg , 1 and 100 μg , 1 and 50 μg , preferably between 1 and 20 μg , and more preferably between 5 and 10 μg of transposon-based vector DNA is administered to the oviduct of a bird. Optimal ranges depend upon the type of bird and the bird's stage of sexual maturity. In a chicken, it is preferred that between approximately 1 and 100 μg , or 5 and 50 μg are administered. In a quail, it is preferred that between approximately 5 and 10 μg are administered. Intraoviduct administration of the transposon-based vectors of the present invention result in incorporation of the gene of interest into the cells of the oviduct as evidenced by a PCR positive signal in the oviduct tissue. In other embodiments, the transposon-based vector is administered to an artery that supplies the oviduct. These methods of administration may also be combined with any methods for facilitating transfection, including without limitation, electroporation, gene guns, injection of naked DNA, and use of dimethyl sulfoxide (DMSO).

[0130] According to the present invention, the transposon-based vector is administered in conjunction with an acceptable carrier and/or transfection reagent. Acceptable carriers include, but are not limited to, water, saline, Hanks Balanced Salt Solution (HBSS), Tris-EDTA (TE) and lyotropic liquid crystals. Transfection reagents commonly known to one of ordinary skill in the art that may be employed include, but are not limited to, the following: cationic lipid transfection reagents, cationic lipid mixtures, polyamine reagents, liposomes and combinations thereof; SUPERFECT®, Cytfectene, BioPORTER®, GenePORTER®, NeuroPORTER®, and perfectin from Gene Therapy Systems; lipofectamine, cellfectin,

DMRIE-C oligofectamine, TROJENE® and PLUS reagent from Invitrogen; Xtreme gene, fugene, DOSPER and DOTAP from Roche; Lipotaxi and Genejammer from Strategene; and Escort from SIGMA. In one embodiment, the transfection reagent is SUPERPECT®. The ratio of DNA to transfection reagent may vary based upon the method of administration. In one embodiment, the transposon-based vector is administered to the oviduct and the ratio of DNA to transfection reagent can be from 1:1.5 to 1:15, preferably 1:2 to 1:5, all expressed as wt/vol. Transfection may also be accomplished using other means known to one of ordinary skill in the art, including without limitation electroporation, gene guns, injection of naked DNA, and use of dimethyl sulfoxide (DMSO).

[0131] Depending upon the cell or tissue type targeted for transfection, the form of the transposon-based vector may be important. Plasmids harvested from bacteria are generally closed circular supercoiled molecules, and this is the preferred state of a vector for gene delivery because of the ease of preparation. In some instances, transposase expression and insertion may be more efficient in a relaxed, closed circular configuration or in a linear configuration. In still other instances, a purified transposase protein may be co-injected with a transposon-based vector containing the gene of interest for more immediate insertion. This could be accomplished by using a transfection reagent complexed with both the purified transposase protein and the transposon-based vector.

Testing for and Breeding Animals Carrying the Transgene

[0132] Following administration of a transposon-based vector to an bird, DNA is extracted from the bird to confirm integration of the gene of interest. Advantages provided by the present invention include the high rates of integration, or incorporation, and transcription of the gene of interest when administered to a bird via an intraoviduct or intraovarian route (including intraarterial administrations to arteries leading to the oviduct or ovary). Example 6 below describes isolation of a proinsulin/ENT TAG protein from a transgenic hen following ammonium sulfate precipitation and ion exchange chromatography. Figure 5 demonstrates successful administration of a transposon-based vector to a hen, successful integration of the gene of interest, successful production of a protein encoded by the gene of interest, and successful deposition of the protein in egg white produced by the transgenic hen.

[0133] Actual frequencies of integration may be estimated both by comparative strength of the PCR signal, and by histological evaluation of the tissues by quantitative PCR. Another method for estimating the rate of transgene insertion is the so-called primed in situ hybridization technique (PRINS). This method determines not only which cells carry a transgene of interest, but also into which chromosome the gene has inserted, and even what portion of the chromosome. Briefly, labeled primers are annealed to chromosome spreads (affixed to glass slides) through one round of PCR, and the slides are then developed through normal in situ hybridization procedures. This technique combines the best features of, in situ PCR and fluorescence in situ hybridization (FISH) to provide distinct chromosome location and copy number of the gene in question.

[0134] Breeding experiments are also conducted to determine if germline transmission of the transgene has occurred. In a general bird breeding experiment performed according to the present invention, each male bird was exposed to 2-3 different adult female birds for 3-4 days each. This procedure was continued with different females for a total period of 6-12 weeks. Eggs are collected daily for up to 14 days after the last exposure to the transgenic male, and each egg is incubated in a standard incubator. The resulting embryos are examined for transgene presence at day 3 or 4 using PCR. It is to be understood that the above procedure can be modified to suit animals other than birds and that selective breeding techniques may be performed to amplify gene copy numbers and protein output.

Production of Desired Proteins or Peptides in Egg White

[0135] In one embodiment, the transposon-based vectors of the present invention may be administered to a bird for production of desired proteins or peptides in the egg white. These transposon-based vectors preferably contain one or more of an ovalbumin promoter, an ovomucoid promoter, an ovalbumin signal sequence and an ovomucoid signal sequence. Oviduct-specific ovalbumin promoters are described in B. O'Malley et al., 1987. EMBO J., vol. 6. pp. 2305-12; A. Qiu et al., 1994. Proc. Nat. Acad. Sci. (USA), vol. 91, pp. 4451-4455; D. Monroe et al., 2000. Biochim. Biophys. Acta, 1517 (1):27-32; H. Park et al., 2000. Biochem., 39:8537-8545; and T. Muramatsu et al., 1996. Poult. Avian Biol. Rev., 6:107-123. Examples of transposon-based vectors designed for production of a desired protein in an egg white are shown in Figures 2 and 3.

Production of Desired Proteins or Peptides in Egg Yolk

[0136] The present invention is particularly advantageous for production of recombinant peptides and proteins of low solubility in the egg yolk. Such proteins include, but are not limited to, membrane-associated or membrane-bound proteins, lipophilic compounds; attachment factors, receptors, and components of second messenger transduction machinery. Low solubility peptides and proteins are particularly challenging to produce using conventional recombinant

protein production techniques (cell and tissue cultures) because they aggregate in water-based, hydrophilic environments. Such aggregation necessitates denaturation and re-folding of the recombinantly-produced proteins, which may deleteriously affect their structure and function. Moreover, even highly soluble recombinant peptides and proteins may precipitate and require denaturation and renaturation when produced in sufficiently high amounts in recombinant protein production systems. The present invention provides an advantageous resolution of the problem of protein and peptide solubility during production of large amounts of recombinant proteins.

[0137] In one embodiment of the present invention wherein germline transfection is obtained via intraovarian administration of the transposon-based vector, deposition of a desired protein into the egg yolk is accomplished in offspring by attaching a sequence encoding a protein capable of binding to the yolk vitellogenin receptor to a gene of interest that encodes a desired protein. This transposon-based vector can be used for the receptor-mediated uptake of the desired protein by the oocytes. In a preferred embodiment, the sequence ensuring the binding to the vitellogenin receptor is a targeting sequence of a vitellogenin protein. The invention encompasses various vitellogenin proteins and their targeting sequences. In a preferred embodiment, a chicken vitellogenin protein targeting sequence is used, however, due to the high degree of conservation among vitellogenin protein sequences and known cross-species reactivity of vitellogenin targeting sequences with their egg-yolk receptors, other vitellogenin targeting sequences can be substituted. One example of a construct for use in the transposon-based vectors of the present invention and for deposition of an insulin protein in an egg yolk is a transposon-based vector containing a vitellogenin promoter, a vitellogenin targeting sequence, a TAG sequence, a pro-insulin sequence and a synthetic polyA sequence. The present invention includes, but is not limited to, vitellogenin targeting sequences residing in the N-terminal domain of vitellogenin, particularly in lipovitellin I. In one embodiment, the vitellogenin targeting sequence contains the polynucleotide sequence of SEQ ID NO:22. In a preferred embodiment, the transposon-based vector contains a transposase gene operably-linked to a constitutive promoter and a gene of interest operably-linked to a liver-specific promoter and a vitellogenin targeting sequence.

Isolation and Purification of Desired Protein or Peptide

[0138] For large-scale production of protein, an bird breeding stock that is homozygous for the transgene is preferred. Such homozygous individuals are obtained and identified through, for example, standard animal breeding procedures or PCR protocols.

[0139] Once expressed, peptides, polypeptides and proteins can be purified according to standard procedures known to one of ordinary skill in the art, including ammonium sulfate precipitation, affinity columns, column chromatography, gel electrophoresis, high performance liquid chromatography, immunoprecipitation and the like. Substantially pure compositions of about 50 to 99% homogeneity are preferred, and 80 to 95% or greater homogeneity are most preferred for use as therapeutic agents.

[0140] In one embodiment of the present invention, the bird in which the desired protein is produced is an egg-laying bird. In a preferred embodiment of the present invention, the animal is an avian and a desired peptide, polypeptide or protein is isolated from an egg white. Egg white containing the exogenous protein or peptide is separated from the yolk and other egg constituents on an industrial scale by any of a variety of methods known in the egg industry. See, e.g., W. Stadelman et al. (Eds.), *Egg Science & Technology*, Haworth Press, Binghamton, NY (1995). Isolation of the exogenous peptide or protein from the other egg white constituents is accomplished by any of a number of polypeptide isolation and purification methods well known to one of ordinary skill in the art. These techniques include, for example, chromatographic methods such as gel permeation, ion exchange, affinity separation, metal chelation, HPLC, and the like, either alone or in combination. Another means that may be used for isolation or purification, either in lieu of or in addition to chromatographic separation methods, includes electrophoresis. Successful isolation and purification is confirmed by standard analytic techniques, including HPLC, mass spectroscopy, and spectrophotometry. These separation methods are often facilitated if the first step in the separation is the removal of the endogenous ovalbumin fraction of egg white, as doing so will reduce the total protein content to be further purified by about 50%.

[0141] To facilitate or enable purification of a desired protein or peptide, transposon-based vectors may include one or more additional epitopes or domains. Such epitopes or domains include DNA sequences encoding enzymatic or chemical cleavage sites including, but not limited to, an enterokinase cleavage site; the glutathione binding domain from glutathione S-transferase; polylysine; hexa-histidine or other cationic amino acids; thioredoxin; hemagglutinin antigen; maltose binding protein; a fragment of gp41 from HIV; and other purification epitopes or domains commonly known to one of skill in the art.

[0142] In one representative embodiment, purification of desired proteins from egg white utilizes the antigenicity of the ovalbumin carrier protein and particular attributes of a TAG linker sequence that spans ovalbumin and the desired protein. The TAG sequence is particularly useful in this process because it contains 1) a highly antigenic epitope, a fragment of gp41 from HIV, allowing for stringent affinity purification, and, 2) a recognition site for the protease enterokinase immediately juxtaposed to the desired protein. In a preferred embodiment, the TAG sequence comprises approximately 50 amino acids. A representative TAG sequence is provided below.

**Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp
Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys
Gly Trp Ile Gly Leu Leu Asp Asp Asp Asp Lys (SEQ ID NO:35)**

The underlined sequences were taken from the hairpin loop domain of HIV gp-41 (SEQ ID NO:33). Sequences in italics represent the cleavage site for enterokinase (SEQ ID NO:34). The spacer sequence upstream of the loop domain was made from repeats of (Pro Ala Asp Asp Ala) (SEQ ID NO:31) to provide free rotation and promote surface availability of the hairpin loop from the ovalbumin carrier protein.

[0143] Isolation and purification of a desired protein is performed as follows:

1. Enrichment of the egg white protein fraction containing ovalbumin and the transgenic ovalbumin-TAG-desired protein.
2. Size exclusion chromatography to isolate only those proteins within a narrow range of molecular weights (a further enrichment of step 1).
3. Ovalbumin affinity chromatography. Highly specific antibodies to ovalbumin will eliminate virtually all extraneous egg white proteins except ovalbumin and the transgenic ovalbumin-TAG-desired protein.
4. gp41 affinity chromatography using anti-gp41 antibodies. Stringent application of this step will result in virtually pure transgenic ovalbumin-TAG-desired protein.
5. Cleavage of the transgene product can be accomplished in at least one of two ways:
 - a. The transgenic ovalbumin-TAG-desired protein is left attached to the gp41 affinity resin (beads) from step 4 and the protease enterokinase is added. This liberates the transgene target protein from the gp41 affinity resin while the ovalbumin-TAG sequence is retained. Separation by centrifugation (in a batch process) or flow through (in a column purification), leaves the desired protein together with enterokinase in solution. Enterokinase is recovered and reused.
 - b. Alternatively, enterokinase is immobilized on resin (beads) by the addition of poly-lysine moieties to a non-catalytic area of the protease. The transgenic ovalbumin-TAG-desired protein eluted from the affinity column of step 4 is then applied to the protease resin. Protease action cleaves the ovalbumin-TAG sequence from the desired protein and leaves both entities in solution. The immobilized enterokinase resin is recharged and reused.
 - c. The choice of these alternatives is made depending upon the size and chemical composition of the transgene target protein.
6. A final separation of either of these two (5a or 5b) protein mixtures is made using size exclusion, or enterokinase affinity chromatography. This step allows for desalting, buffer exchange and/or polishing, as needed.

[0144] Cleavage of the transgene product (ovalbumin-TAG-desired protein) by enterokinase, then, results in two products: ovalbumin-TAG and the desired protein. More specific methods for isolation using the TAG label is provided in the Examples. Some desired proteins may require additions or modifications of the above-described approach as known to one of ordinary skill in the art. The method is scaleable from the laboratory bench to pilot and production facility largely because the techniques applied are well documented in each of these settings.

[0145] In another representative embodiment, egg whites containing a protein of interest were pooled and separated, in any order, from the yolks and other egg constituents by methods known to one skilled in the art. A variety of such methods is described in manuals known in the art, such as *Egg Science & Technology*, W. Stadelman, et al. (Eds.), Haworth Press, Binghamton, NY (1995).

[0146] One non-limiting example of a method for isolating a desired peptide, polypeptide or protein from an egg white is as follows. It is to be understood that this method may be employed to isolate any desired peptide, polypeptide or protein from the eggs of transgenic animals of the present invention. This present example involved transgenes that used a portion of or the entire Ovalbumin protein, or specific ovalbumin epitopes, as a carrier, linked to the protein of interest via the specified TAG sequence, or another affinity/cleavage sequence. The TAG sequence contains the hairpin loop epitope from HIV I followed by an enterokinase cleavage site.

[0147] First, the viscosity of the egg white was lowered by subjecting the egg white to low shear forces of 3140 cps (Tung et al., 1969). The resulting pourable solution was then filtered to remove chalazae. An ammonium sulfate precipitation was then used to enrich the fraction of transgenic protein (see, for example, *Practical Protein Chemistry A Handbook* A. Darbre (Ed.), John Wiley & Sons Ltd, 1986). Other methods of crude fractionation known in the art are also used as needed. The supernatant of this separation was then fractionated using size-exclusion chromatography, further enriching

the transgenic fusion protein fraction and eliminating the ammonium sulfate from the material. The fusion protein was isolated by anti-ovalbumin affinity chromatography (batch or column) using methods known to one skilled in the art. This step may capture native ovalbumin in addition to an ovalbumin-transgene fusion protein. After elution from the anti-ovalbumin affinity resin, the transgenic protein was specifically isolated using anti-gp41 affinity chromatography (batch or column) using methods known to one skilled in the art.

[0148] Cleavage of the transgene product from the carrier and the TAG sequences was accomplished in one of at least two ways:

1) The transgenic ovalbumin-TAG-transgene target protein was left attached to the gp41 affinity resin and the protease enterokinase was added. Cleavage of the transgene by enterokinase liberated the transgene target protein from the gp41 affinity resin while the ovalbumin-TAG sequence was retained. Separation by centrifugation (in a batch process) or flow through (in a column purification), kept the transgene target protein together with enterokinase in solution. Enterokinase was recovered and reused.

2) Alternatively, enterokinase was immobilized on resin (beads) by the addition of poly-lysine moieties to a non-catalytic area of the protease. The transgenic ovalbumin-TAG-transgene target protein was eluted from the gp41 affinity chromatography resin and then applied to the protease resin. Protease action cleaved the ovalbumin-TAG sequence from the transgene target protein and left both entities in solution. The immobilized enterokinase resin was recharged and reused. The choice between these alternatives is made on a case-by case basis, depending upon the size and chemical composition of the transgene target protein.

[0149] A final separation of either of these two (process or 2) protein mixtures was made using size exclusion chromatography, or enterokinase affinity chromatography. This step also allows for desalting, concentrating, buffer exchange and/or polishing, as needed.

[0150] It is believed that a typical chicken egg produced by a transgenic animal of the present invention will contain at least 0.001 mg, from about 0.001 to 1.0 mg, or from about 0.001 to 100.0 mg of exogenous protein, peptide or polypeptide, in addition to the normal constituents of egg white (or possibly replacing a small fraction of the latter). In some embodiments, a chicken egg will contain between 50 and 75 mg of exogenous protein.

[0151] One of skill in the art will recognize that after biological expression or purification, the desired proteins, fragments thereof and peptides may possess a conformation substantially different than the native conformations of the proteins, fragments thereof and peptides. In this case, it is often necessary to denature and reduce protein and then to cause the protein to re-fold into the preferred conformation. Methods of reducing and denaturing proteins and inducing re-folding are well known to those of skill in the art.

Production of Protein or Peptide in Milk

[0152] In addition to methods of producing eggs containing transgenic proteins or peptides, the present invention makes reference for comparison to methods for the production of milk containing transgenic proteins or peptides. These methods include the administration of a transposon-based vector described above to a mammal through the duct system.

[0153] The transposon-based vector may contain a transposase operably-linked to a constitutive promoter and a gene of interest operably-linked to mammary specific promoter. Genes of interest can include, but are not limited to antiviral and antibacterial proteins and immunoglobulins. Further, a transposon-based vector may be administered to the ovary of an animal and germline transformation is obtained. In such cases, offspring of the transfected animal express a gene of interest in the mammary gland under the control of a mammary gland-specific promoter.

[0154] The following examples will serve to further illustrate the present invention without, at the same time, however, constituting any limitation thereof. On the contrary, it is to be clearly understood that resort may be had to various embodiments, modifications and equivalents thereof which, after reading the description herein, may suggest themselves to those skilled in the art without departing from the spirit of the invention.

EXAMPLE 1

IntraOviduct Administration of Transposon-Based Vectors

[0155] Quail or chicken were selected for administration of the transposon-based vectors of the present invention. Feathers were removed from the area where surgery was performed and the area was cleansed and sterilized by rinsing it with ethanol (alcohol) and 0.5% chlorhexidine. Using the scalpel, a dorsolateral incision was made through the skin over the ovary approximately 2 cm in length. Using blunt scissors, a second incision was made through the muscle between the last two ribs to expose the oviduct beneath. A small animal retractor was used to spread the last two ribs,

exposing the oviduct beneath. The oviduct was further exposed using retractors to pull the intestines to one side.

[0156] A delivery solution containing a transposon-based vector and SUPERFECT® was prepared fresh immediately before surgery. Specific ratios of vector and SUPERFECT® that were used in each experiment are provided in the Examples below. The delivery solution was warmed to room temperature prior to injection into the bird. Approximately 250-500 µl of the delivery solution was injected into the lumen of the magnum of the oviduct using a 1 cc syringe with a 27 gauge needle attached. The wound was closed and antibiotic cream liberally applied to the area surrounding the wound.

EXAMPLE 2

Preparation of Transposon-Based Vector pTnMod

[0157] A vector was designed for inserting a desired coding sequence into the genome of eukaryotic cells, given below as SEQ ID NO:3. The vector of SEQ ID NO:3, termed pTnMod, was constructed and its sequence verified.

[0158] This vector employed a cytomegalovirus (CMV) promoter. A modified Kozak sequence (ACCATG) (SEQ ID NO:1) was added to the promoter. The nucleotides in the wobble position in nucleotide triplet codons encoding the first 10 amino acids of transposase was changed to an adenine (A) or thymine (T), which did not alter the amino acid encoded by this codon. Two stop codons were added and a synthetic polyA was used to provide a strong termination sequence. This vector uses a promoter designed to be active soon after entering the cell (without any induction) to increase the likelihood of stable integration. The additional stop codons and synthetic polyA insures proper termination without read through to potential genes downstream.

[0159] The first step in constructing this vector was to modify the transposase to have the desired changes. Modifications to the transposase were accomplished with the primers High Efficiency forward primer (Hef) Altered transposase (ATS)-Hef 5' ATCTCGAGACCATGTGTGAACCTTGATATTTTACATGATTCTCTTTACC 3' (SEQ ID NO:36) and Altered transposase- High efficiency reverse primer (Her) 5' GATTGATCATTATCATAATTTCCCAAAGCGTAACC 3' (SEQ ID NO:37, a reverse complement primer). In the 5' forward primer ATS-Hef, the sequence CTCGAG (SEQ ID NO:38) is the recognition site for the restriction enzyme Xho I, which permits directional cloning of the amplified gene. The sequence ACCATG (SEQ ID NO:1) contains the Kozak sequence and start codon for the transposase and the underlined bases represent changes in the wobble position to an A or T of codons for the first 10 amino acids (without changing the amino acid coded by the codon). Primer ATS-Her (SEQ ID NO:37) contains an additional stop codon TAA in addition to native stop codon TGA and adds a Bcl I restriction site, TGATCA (SEQ ID NO:39), to allow directional cloning. These primers were used in a PCR reaction with pTnLac (p defines plasmid, tn defines transposon, and lac defines the beta fragment of the lactose gene, which contains a multiple cloning site) as the template for the transposase and a FailSafe™ PCR System (which includes enzyme, buffers, dNTP's, MgCl₂ and PCR Enhancer; Epicentre Technologies, Madison, WI). Amplified PCR product was electrophoresed on a 1% agarose gel, stained with ethidium bromide, and visualized on an ultraviolet transilluminator. A band corresponding to the expected size was excised from the gel and purified from the agarose using a Zymo Clean Gel Recovery Kit (Zymo Research, Orange, CA). Purified DNA was digested with restriction enzymes Xho I (5') and Bcl I (3') (New England Biolabs, Beverly, MA) according to the manufacturer's protocol. Digested DNA was purified from restriction enzymes using a Zymo DNA Clean and Concentrator kit (Zymo Research).

[0160] Plasmid gWhiz (Gene Therapy Systems, San Diego, CA) was digested with restriction enzymes Sal I and BamH I (New England Biolabs), which are compatibly with Xho I and Bcl I, but destroy the restriction sites. Digested gWhiz was separated on an agarose gel, the desired band excised and purified as described above. Cutting the vector in this manner facilitated directional cloning of the modified transposase (mATS) between the CMV promoter and synthetic polyA.

[0161] To insert the mATS between the CMV promoter and synthetic polyA in gWhiz, a Stratagene T4 Ligase Kit (Stratagene, Inc. La Jolla, CA) was used and the ligation set up according to the manufacturer's protocol. Ligated product was transformed into *E. coli* Top10 competent cells (Invitrogen Life Technologies, Carlsbad, CA) using chemical transformation according to Invitrogen's protocol. Transformed bacteria were incubated in 1 ml of SOC (GIBCO BRL, CAT# 15544-042) medium for 1 hour at 37° C before being spread to LB (Luria-Bertani media (broth or agar)) plates supplemented with 100 µg/ml ampicillin (LB/amp plates). These plates were incubated overnight at 37° C and resulting colonies picked to LB/amp broth for overnight growth at 37° C. Plasmid DNA was isolated using a modified alkaline lysis protocol (Sambrook et al., 1989), electrophoresed on a 1% agarose gel, and visualized on a U.V. transilluminator after ethidium bromide staining. Colonies producing a plasmid of the expected size (approximately 6.4 kbp) were cultured in at least 250 ml of LB/amp broth and plasmid DNA harvested using a Qiagen Maxi-Prep Kit (column purification) according to the manufacturer's protocol (Qiagen, Inc., Chatsworth, CA). Column purified DNA was used as template for sequencing to verify the changes made in the transposase were the desired changes and no further changes or mutations occurred due to PCR amplification. For sequencing, Perkin-Elmer's Big Dye Sequencing Kit was used. All samples were sent to the Gene Probes and Expression Laboratory (LSU School of Veterinary Medicine) for sequencing on a Perkin-Elmer

Model 377 Automated Sequencer..

[0162] Once a clone was identified that contained the desired mATS in the correct orientation, primers CMVf-NgoM IV (5' TTGCCGGCATCAGATTGGCTAT (SBQ ID NO:40); underlined bases denote a NgoM IV recognition site) and Syn-polyA-BstE II (5' AGAGGTCACCGGGTCAATTCTTCAGCACCTGGTA (SEQ ID NO:41); underlined bases denote a BstE II recognition site) were used to PCR amplify the entire CMV promoter, mATS, and synthetic polyA for cloning upstream of the transposon in pTnLac. The PCR was conducted with FailSafe™ as described above, purified using the Zymo Clean and Concentrator kit, the ends digested with NgoM IV and BstE II (New England Biolabs), purified with the Zymo kit again and cloned upstream of the transposon in pTnLac as described below.

[0163] Plasmid pTnLac was digested with NgoM IV and BstE II to remove the ptac promoter and transposase and the fragments separated on an agarose gel. The band corresponding to the vector and transposon was excised, purified from the agarose, and dephosphorylated with calf intestinal alkaline phosphatase (New England Biolabs) to prevent self-annealing. The enzyme was removed from the vector using a Zymo DNA Clean and Concentrator-5. The purified vector and CMVp/mATS/polyA were ligated together using a Stratagene T4 Ligase Kit and transformed into *E. coli* as described above.

[0164] Colonies resulting from this transformation were screened (mini-preps) as describe above and clones that were the correct size were verified by DNA sequence analysis as described above. The vector was given the name pTnMod (SEQ ID NO:3) and includes the following components:

Base pairs 1-130 are a remainder of F1(-) on from pBluescriptII sk(-) (Stratagene), corresponding to base pairs 1-130 of pBluescriptII sk(-).

Base pairs 131 - 132 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 133 -1777 are the CMV promoter/enhancer taken from vector pGWiz (Gene Therapy Systems), corresponding to bp 229-1873 of pGWiz. The 'CMV promoter was modified by the addition of an ACC sequence upstream of ATG.

Base pairs 1778-1779 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 1780 - 2987 are the coding sequence for the transposase, modified from Tn10 (GenBank accession J01829) by optimizing codons for stability of the transposase mRNA and for the expression of protein. More specifically, in each of the codons for the first ten amino acids of the transposase, G or C was changed to A or T when such a substitution would not alter the amino acid that was encoded.

Base pairs 2988-2993 are two engineered stop codons.

Base pair 2994 is a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 2995 - 3410 are a synthetic polyA sequence taken from the pGWiz vector (Gene Therapy Systems), corresponding to bp 1922-2337 of 10 pGWiz.

Base pairs 3415 - 3718 are non-coding DNA that is residual from vector pNK2859.

Base pairs 3719 - 3761 are non-coding λ DNA that is residual from pNK2859.

Base pairs 3762 - 3831 are the 70 bp of the left insertion sequence recognized by the transposon Tn10.

Base pairs 3832-3837 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 3838 - 4527 are the multiple cloning site from pBluescriptII sk(20), corresponding to bp 924-235 of pBluescriptII sk(-). This multiple cloning site may be used to insert any coding sequence of interest into the vector.

Base pairs 4528-4532 are a residue from ligation of restriction enzyme sites used in constructing the vector.

Base pairs 4533 - 4602 are the 70 bp of the right insertion sequence recognized by the transposon Tn10.

Base pairs 4603 - 4644 are non-coding λ DNA that is residual from pNK2859.

Base pair 4645 - 5488 are non-coding DNA that is residual from pNK2859.

Base pairs 5489 - 7689 are from the pBluescriptII sk(-) base vector - (Stratagene, Inc.), corresponding to bp 761-2961 of pBluescriptII sk(-).

[0165] Completing pTnMod is a pBlueScript backbone that contains a colE I origin of replication and an antibiotic resistance marker (ampicillin).

[0166] It should be noted that all non-coding DNA sequences described above can be replaced with any other non-coding DNA sequence(s). Missing nucleotide sequences in the above construct represent restriction site remnants.

[0167] All plasmid DNA was isolated by standard procedures. Briefly, *Escherichia coli* containing the plasmid was grown in 500 mL aliquots of LB broth (supplemented with an appropriate antibiotic) at 37°C overnight with shaking. Plasmid DNA was recovered from the bacteria using a Qiagen Maxi-Prep kit (Qiagen, Inc., Chatsworth, CA) according to the manufacturer's protocol. Plasmid DNA was resuspended in 500 μ L of PCR-grade water and stored at -20°C until used.

EXAMPLE 3

Transposon-Based Vector pTnMCS

5 **[0168]** Another transposon-based vector was designed for inserting a desired coding sequence into the genome of eukaryotic cells. This vector was termed pTnMCS and its constituents are provided below. The sequence of the pTnMCS vector is provided in SEQ ID NO:2. The pTnMCS vector contains an avian optimized polyA sequence operably-linked to the transposase gene. The avian optimized polyA sequence contains approximately 40 nucleotides that precede the A nucleotide string.

10 Bp 1-130 Remainder of F1 (-) ori of pBluescriptII sk(-) (Stratagene) bp1-130
 Bp 133 - 1777 CMV promoter/enhancer taken from vector pGWIZ (Gene Therapy Systems) bp 229-1873
 Bp 1783 - 2991 Transposase, from Tn10 (GenBank accession #J01829) bp 108-1316
 Bp 2992 - 3344 Non coding DNA from vector pNK2859
 15 Bp 3345 - 3387 Lambda DNA from pNK2859
 Bp 3388 - 3457 70 bp of IS10 left from Tn10
 Bp 3464 - 3670 Multiple cloning site from pBluescriptII sk(-), thru the XmaI site bp 924-718
 Bp 3671 - 3715 Multiple cloning site from pBluescriptII sk(-), from the XmaI site thru the XhoI site. These base pairs are usually lost when cloning into pTnMCS bp 717-673
 20 Bp 3716 - 4153 Multiple cloning site from pBluescriptII sk(-), from the XhoI site bp 672-235
 Bp 4159 - 4228 70 bp of IS10 right from Tn10
 Bp 4229 - 4270 Lambda DNA from pNK2859
 Bp 4271 - 5114 Non-coding DNA from pNK2859
 Bp 5115 - 7315 pBluescript sk (-) base vector (Stratagene, Inc.) bp 761-2961.

EXAMPLE 4

Preparation of Transposon-Based Vector pThMod(Oval/ENT TAG/ProIns/PA)-Chicken

30 **[0169]** A vector was designed to insert a human proinsulin coding sequence under the control of a chicken ovalbumin promoter, and a ovalbumin gene including an ovalbumin signal sequence, into the genome of a bird given below as SEQ ID NO:42.

35 Base pairs 1 - 130 are a remainder of F1(-) ori of pBluescriptII sk(-) (Stratagene) corresponding to base pairs 1-130 of pBluescriptII sk(-).
 Base pairs 133 - 1777 are a CMV promoter/enhancer taken from vector pGWiz (Gene Therapy Systems) corresponding to base pairs 229-1873 of pGWiz.
 Base pairs 1780 - 2987 are a transposase, modified from Tn10 (GenBank accession number J01829).
 Base pairs 2988-2993 are two engineered stop codons.
 40 Base pairs 2995 - 3410 are a synthetic polyA from pGWiz (Gene Therapy Systems) corresponding to base pairs 1922- 2337 of pGWiz.
 Base pairs 3415 - 3718 are non coding DNA that is residual from vector pNK2859.
 Base pairs 3719 - 3761 are X DNA that is residual from pNK2859.
 Base pairs 3762 - 3831 are the 70 base pairs of the left insertion sequence (IS10) recognized by the transposon Tn10.
 45 Base pairs 3838 - 4044 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 924-718 of pBluescriptII sk(-).
 Base pairs 4050 - 4951 are a chicken ovalbumin promoter (including SDRE) that corresponds to base pairs 431-1332 of the chicken ovalbumin promoter in GenBank Accession Number J00895 M24999.
 Base pairs 4958 - 6115 are a chicken ovalbumin signal sequence and ovalbumin gene that correspond to base pairs 66-1223 of GenBank Accession Number V00383.1. (The STOP codon being omitted).
 50 Base pairs 6122 - 6271 are a TAG. sequence containing a gp41 hairpin loop from HIV I, an enterokinase cleavage site and a spacer (synthetic).
 Base pairs 6272 - 6531 are a proinsulin gene.
 Base pairs 6539 - 6891 are a synthetic polyadenylation sequence from pGWiz (Gene Therapy Systems) corresponding to base pairs 1920 - 2272 of pGWiz.
 55 Base pairs 6897 - 7329 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 667-235 of pBluescriptII sk(-).
 Base pairs 7335- 7404 are the 70 base pairs of the right insertion sequence (IS10) recognized by the transposon Tn10.

EP 1 592 789 B1

Base pairs 7405 - 7446 are λ DNA that is residual from pNK2859.

Base pairs 7447 - 8311 are non coding DNA that is residual from pNK2859.

Base pairs 8312 - 10512 are pBlueScript sk(-) base vector (Stratagene, Inc.) corresponding to base pairs 761-2961 of pBluescriptII sk(-).

5

[0170] It should be noted that all non-coding DNA sequences described above can be replaced with any other non-coding DNA sequence(s). Missing nucleotide sequences in the above construct represent restriction site remnants.

EXAMPLE 5

10

Transposon-Based Vector pTnMOD (CMV-CHOVg-ent-Proinsulin-synPA)

[0171] A vector was designed to insert a proinsulin coding sequence under the control of a quail ovalbumin promoter, and a ovalbumin gene including an ovalbumin signal sequence, into the genome of a bird given below as SEQ ID NO:43.

15

Bp 1- 4045 from vector pTnMod, bp 1- 4045

Bp 4051- 5695 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864

Bp 5702 -6855 Chicken ovalbumin gene taken from GenBank accession # V00383, bp 66-1219

Bp 6862 - 7011 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site

20

Bp 7012 - 7272 Human Proinsulin taken from GenBank accession # NM000207, bp 117-377

Bp 7273 - 7317 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and pGWIZ (Gene Therapy Systems)

Bp 7318 - 7670 Synthetic polyA from the cloning vector pGWIZ (Gene Therapy Systems), bp 1920-2271

Bp 7672-11271 from cloning vector pTnMCS, bp 3716-7315

25

EXAMPLE 6

Transfection of Japanese Quail using a Transposon-based Vector containing a Proinsulin Gene via Oviduct Injections

30

[0172] Two experiments were conducted in Japanese quail using transposon-based vectors containing either Oval promoter/Oval gene/GP41 Enterokinase TAG/Proinsulin/Poly A (SEQ ID NO:42) or CMV promoter/Oval gene/GP41 Enterokinase TAG/Proinsulin/Poly A (SEQ ID NO:43).

35

[0173] In the first experiment, the Oval promoter/Oval gene/GP41 Enterokinase TAG/Proinsulin/Poly A containing construct was injected into the lumen of the oviduct of sexually mature quail; three hens received 5 μ g at a 1:3 SUPERFECT® ratio and three received 10 μ g at a 1:3 SUPERFECT® ratio. As of the writing of the present application, at least one bird that received above-mentioned construct was producing human proinsulin in egg white (other birds remain to be tested). This experiment indicates that 1) the DNA has been stable for at least 3 months; 2) protein levels are comparable to those observed with a constitutive promoter such as the CMV promoter, and 3) sexually mature birds can be injected and results obtained without the need for cell culture. It is estimated that each quail egg contains approximately 1.4 μ g/ml of the proinsulin protein. It is also estimated that each transgenic chicken egg contains 50-75 mg of protein encoded by the gene of interest

40

[0174] In the second experiment, the transposon-based vector containing CMV promoter/Oval gene/GP41 Enterokinase TAG/Proinsulin/Poly A was injected into the lumen of the oviduct of sexually immature Japanese quail. A total of 9 birds were injected. Of the 8 survivors, 3 produced human proinsulin in the white of their eggs for over 6 weeks. An ELISA assay described in detail below was developed to detect GP41 in the fusion peptide (Oval gene/GP41 Enterokinase TAG/Proinsulin) since the GP41 peptide sequence is unique and not found as part of normal egg white protein. In all ELISA assays, the same birds produced positive results and all controls worked as expected.

45

[0175] ELISA Procedure: Individual egg white samples were diluted in sodium carbonate buffer, pH 9.6, and added to individual wells of 96 well microtiter ELISA plates at a total volume of 0.1 ml. These plates were then allowed to coat overnight at 4°C. Prior to ELISA development, the plates were allowed warm to room temperature. Upon decanting the coating solutions and blotting away any excess, non-specific binding of antibodies was blocked by adding a solution of phosphate buffered saline (PBS), 1% (w/v) BSA, and 0.05% (v/v) Tween 20 and allowing it to incubate with shaking for a minimum of 45 minutes. This blocking solution was subsequently decanted and replaced with a solution of the primary antibody (Goat Anti-GP41 TAG) diluted in fresh PBS/BSA/Tween 20. After a two hour period of incubation with the primary antibody, each plate was washed with a solution of PBS and 0.05% Tween 20 in an automated plate washer to remove unbound antibody. Next, the secondary antibody, Rabbit anti-Goat Alkaline Phosphatase-conjugated, was diluted in PBS/BSA/Tween 20 and allowed to incubate 1 hour. The plates were then subjected to a second wash with PBS/Tween 20. Antigen was detected using a solution of p-Nitrophenyl Phosphate in Diethanolamine Substrate Buffer for

55

Alkaline Phosphatase and measuring the absorbance at 30 minutes and 1 hour.

[0176] Additionally, a proinsulin fusion protein produced using a construct described above was isolated from egg white using ammonium sulfate precipitation and ion exchange, chromatography. A pooled fraction of the isolated fusion protein was run on an SDS-PAGE gel shown in Figure 5, lanes 4 and 6. Lanes 1 and 10 of the gel contain molecular weight standards, lanes 2 and 8 contain non-transgenic chicken egg white, whereas lanes 3, 5, 7 and 9 are blank.

EXAMPLE 7

Isolation of Human Proinsulin Using Anti-TAG Column Chromatography

[0177] A HiTrap NHS-activated 1 mL column (Amersham) was charged with a 30 amino acid peptide that contained the gp-41 epitope containing gp-41's native disulfide bond that stabilizes the formation of the gp-41 hairpin loop. The 30 amino acid gp41 peptide is provided as SEQ ID NO:32. Approximately 10 mg of the peptide was dissolved in coupling buffer (0.2 M NaHCO₃, 0.5 M NaCl, pH 8.3 and the ligand was circulated on the column for 2 hours at room temperature at 0.5 mL/minute. Excess active groups were then deactivated using 6 column volumes of 0.5 M ethanolamine, 0.5 M NaCl, pH 8.3 and the column was washed alternately with 6 column volumes of acetate buffer (0.1 M acetate, 0.5 M NaCl, pH 4.0) and ethanolamine (above). The column was neutralized using 1 X PBS. The column was then washed with buffers to be used in affinity purification: 75 mM Tris, pH 8.0 and elution buffer, 100 mM glycine-HCl, 0.5 M NaCl, pH 2.7. Finally, the column was equilibrated in 75 mM Tris buffer, pH 8.0.

[0178] Antibodies to gp-41 were raised in goats by inoculation with the gp-41 peptide described above. More specifically, goats were inoculated, given a booster injection of the gp-41 peptide and blood samples were obtained by veinupuncture. Serum was harvested by centrifugation. Approximately 30 mL of goat serum was filtered to 0.45 μ M and passed over a TAG column at a rate of 0.5 mL/min. The column was washed with 75 mM Tris, pH 8.0 until absorbance at 280 nm reached a baseline. Three column volumes (3 mL) of elution buffer (100 mM glycine, 0.5 M NaCl, pH 2.7) was applied, followed by 75 mM Tris buffer, pH 8.0, all at a rate of 0.5 mL/min. One milliliter fractions were collected. Fractions were collected into 200 μ L 1 M Tris, pH 9.0 to neutralize acidic fractions as rapidly as possible. A large peak eluted from the column, coincident with the application the elution buffer. Fractions were pooled. Analysis by SDS-PAGE showed a high molecular weight species that separated into two fragments under reducing condition, in keeping with the heavy and light chain structure of IgG.

[0179] Pooled antibody fractions were used to charge two 1 mL HiTrap NHS-activated columns, attached in series. Coupling was carried out in the same manner as that used for charging the TAG column.

Isolation of Ovalbumin- TAG-Pro insulin from Egg White

[0180] Egg white from quail and chickens treated by intra-oviduct injection of the CMV-ovalbumin-TAG-proinsulin construct were pooled. Viscosity was lowered by subjecting the allantoic fluid to successively finer pore sizes using negative pressure filtration, finishing with a 0.22 μ M pore size. Through the process, egg white was diluted approximately 1:16. The clarified sample was loaded on the Anti-TAG column and eluted in the same manner as described for the purification of the anti-TAG antibodies. A peak of absorbance at 280 nm, coincident with the application of the elution buffer, indicated that protein had been specifically eluted from the Anti-TAG column. Fractions containing the eluted peak were pooled for analysis.

[0181] The pooled fractions from the Anti-TAG affinity column were characterized by SDS-PAGE and western blot analysis. SDS-PAGE of the pooled fractions revealed a 60 kDa molecular weight band not present in control egg white fluid, consistent with the predicted molecular weight of the transgenic protein. Although some contaminating bands were observed, the 60 kDa species was greatly enriched compared to the other proteins. An aliquot of the pooled fractions was cleaved overnight at room temperature with the protease, enterokinase. SDS-PAGE analysis of the cleavage product, revealed a band not present in the uncut material that co-migrated with a commercial human proinsulin positive control. Western blot analysis showed specific binding to the 60 kDa species under non-reducing condition (which preserved the hairpin epitope of gp-41 by retaining the disulfide bond). Western analysis of the low molecular weight species that appeared upon cleavage with an anti-human proinsulin antibody, conclusively identified the cleaved fragment as human proinsulin.

EXAMPLE 8

Purification Procedures for Insulin

[0182] L ELISA data for egg characterization/identification

[0183] An ELISA was employed for the initial screening of eggs and, thereby, identification of hens producing positive

eggs. With further modifications this procedure was used for the initial quantification of recombinant protein amounts. These procedures were aided by the successful purification of an initial stock of the recombinant proinsulin (RPI). This stock of protein is used in the development of a double antibody assay that increases the sensitivity and reduces the background in the assay. Subsequent identification of hens producing positive eggs obviate the need to screen each egg collected. Only periodic checks are needed to determine if production levels are consistent

II. Egg White (EW) or Albumin Preparation

A. Clarification - Ovomucin precipitation

[0184] Eggs from hens positively identified as producing RPI are pooled for RPI purification. The initial purification step involved diluting the pool 1:1 with 100 mM Tris-HCl, pH 8 for a final concentration of 50 mM Tris-HCl. The pH of this solution was then adjusted to 6 and ovomucin was allowed to precipitate at 4°C for a minimum of 3hrs (preferably overnight) with constant stirring. The precipitated ovomucin was then pelleted and removed by centrifugation at 2400 x g. After collection of the RPI containing supernatant, the pH of this solution was readjusted to 8.

B. Filtration

[0185] To prepare the egg white for loading onto the column and, thereby, minimize the potential for clogging the columns during loading, the egg white solution was filtered to at least 0.45 μ m.

[0186] Initially, the ovomucin precipitated egg white solution was subjected to successive filtration steps with the pore size of the filtration membrane decreasing at each step. This procedure involved time and dilution of the egg white solution to reach 0.45 μ m filtration.

[0187] Amersham's hollow-fiber ultrafiltration apparatus was used to produce a column-ready solution filtered down to < 0.2 μ m with an undiluted starting solution. This approach minimized the time and the solution dilution needed to prepare the egg white solution for column loading.

III. Purification

A. Affinity Chromatography

[0188] Using antibody with specificity to a synthetic peptide modeled after the enterokinase recognition site, initial purification schemes involved developing a one-step column purification procedure for the RPI.

[0189] Goats immunized with the synthetic Ent peptide were employed to produce anti-Ent Tag antiserum which was used in the egg screening ELISAs followed by antibody purification. The purified goat Anti-Ent Tag antibodies were covalently bound to the matrix of HiTrap NHS-activated HPL columns (Amersham) and subsequently used to specifically bind and purify the RPI.

[0190] An initial attempt was made to direct the first purification step against the ovalbumin portion of the recombinant protein using an antibody specific for the ovalbumin portion. The present purification scheme employed a combination of classical techniques such as ammonium sulfate precipitation, ion exchange, and gel filtration chromatography.

[0191] After the initial ovomucin precipitation, the egg white solution was subjected to protein precipitation using a 40% ammonium sulfate fractionation. The precipitated protein was subsequently collected via centrifugation and resuspended in 50 mM Tris-HCl, pH 8. The resuspended protein solution was dialyzed to remove residual $(\text{NH}_4)_2\text{SO}_4$ or subjected to gel filtration to remove the $(\text{NH}_4)_2\text{SO}_4$ and partially isolate the RPI from the remaining egg white protein. The RPI was further isolated via anion exchange chromatography using a 0 to 0.5M NaCl gradient in 50 mM Tris-HCl, pH 8. Two possible elution profiles were observed. One at approximately 25% of the 0.5 M NaCl gradient without $(\text{NH}_4)_2\text{SO}_4$ precipitation. The second was observed at less than 16% gradient (approximately 7%) following 40% $(\text{NH}_4)_2\text{SO}_4$ precipitation and a longer gradient. Fractions containing RPI were identified by SDS-PAGE analysis and pooled.

[0192] Three gel filtration columns, differing by column size and fractionation range, were employed in RPI purification and/or desalting. Superdex 75 10/300 GL, Hiload 26/60 Superdex 75, and Hiload 26/60 Superdex 200. Using these individual columns at different steps in the purification scheme increased the efficiency of the process. Fractions containing RPI were identified by SDS-PAGE analysis and pooled.

[0193] Cleavage of the RPI Enterokinase recognition site was accomplished using purified enterokinase from Sigma. Enterokinase, 0.004 Unit/ μ l per reaction, was applied to the pooled and, if necessary, concentrated protein solution. The digestion reaction was incubated at room temperature (up to 30°C in a rolling hybridization oven) for a minimum of 16 h and in some cases up to 48 hrs of incubation. The digestion efficiency was followed using 16.5% Tris-Tricine SDS-PAGE peptide gels. All gel staining utilized Simply Blue Coomassie Staining Solutions. Free Proinsulin was observed

on gels after digestion.

[0194] A subsequent gel filtration separation was employed to obtain purified Proinsulin, and to remove the remaining Ovalbumin portion of the RPI and residual native EW proteins. Select steps in the purification process were analyzed using the 2-dimensional Beckman Coulter ProteomeLab PF2D Protein Fractionation System.

EXAMPLE 9

Optimization of Intra-oviduct and Intra-ovarian Arterial Injections

[0195] Overall transfection rates of oviduct cells in a flock of chicken or quail hens are enhanced by synchronizing the development of the oviduct and ovary within the flock. When the development of the oviducts and ovaries are uniform across a group of hens and when the stage of oviduct and ovarian development can be determined or predicted, timing of injections is optimized to transfect the greatest number of cells. Accordingly, oviduct development is synchronized as described below to ensure that a large and uniform proportion of oviduct secretory cells are transfected with the gene of interest.

[0196] Hens are treated with estradiol to stimulate oviduct maturation as described in Oka and Schimke (T. Oka and RT Schimke, J. Cell Biol., 41, 816 (1969)), Palmiter, Christensen and Schimke (J Biol. Chem. 245(4):833-845, 1970). Specifically, repeater daily injections of 1 mg estradiol benzoate are performed sometime before the onset of sexual maturation, a period ranging from 1 - 14 weeks of age. After a stimulation period sufficient to maximize development of the oviduct, hormone treatment is withdrawn thereby causing regression in oviduct secretory cell size but not cell number. At an optimum time after hormone withdrawal, the lumens of the oviducts of treated hens are injected with the transposon-based vector. Hens are subjected to additional estrogen stimulation after an optimized time during which the transposon-based vector is taken up into oviduct secretory cells. Re-stimulation by estrogen activates transposon expression, causing the integration of the gene of interest into the host genome. Estrogen stimulation is then withdrawn and hens continue normal sexual development. If a developmentally regulated promoter such as the ovalbumin promoter is used, expression of the transposon-based vector initiates in the oviduct at the time of sexual maturation. Intra-ovarian artery injection during this window allows for high and uniform transfection efficiencies of ovarian follicles to produce germ-line transfections and possibly oviduct expression.

[0197] Other means are also used to synchronize the development, or regression, of the oviduct and ovary to allow high and uniform transfection efficiencies. Alterations of lighting and/or feed regimens, for example, cause hens to 'molt' during which time the oviduct and ovary regress. Molting is used to synchronize hens for transfection, and may be used in conjunction with other hormonal methods to control regression and/or development of the oviduct and ovary.

EXAMPLE 10

Preparation of Trisposon-Based Vector pTnMod(oval/ENT TAG/ProIns/PA)-Quail

[0198] A vector is designed for inserting a proinsulin gene under the control of a quail ovalbumin promoter, and a ovalbumin gene including an ovalbumin signal sequence, into the genome of a bird given below as SEQ ID NO:44.

Base pairs 1 -130 are a remainder of F1(-) ori of pBluescriptII sk(-) (Stratagene) corresponding to base pairs 1-130 of pBluescriptII sk(-).

Base pairs 133 - 1777 are a CMV promoter/enhancer taken from vector pGWiz (Gene Therapy Systems) corresponding to base pairs 229-1873 of pGWiz.

Base pairs 1780 - 2987 are a transposase, modified from Tn10 (GenBank accession number J01829).

Base pairs 2988-2993 are an engineered stop codon.

Base pairs 2995 - 3410 are a synthetic polyA from pGWiz (Gene Therapy Systems) corresponding to base pairs 1922- 2337 of pGWiz.

Base pairs 3415 - 3718 are non coding DNA that is residual from vector pNK2859.

Base pairs 3719 - 3761 are λ DNA that is residual from pNK2859.

Base pairs 3762 - 3831 are the 70 base pairs of the left insertion sequence (IS10) recognized by the transposon Tn10.

Base pairs 3838 - 4044 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 924-718 of pBluescriptII sk(-).

Base pairs 4050 - 4938 are the Japanese quail ovalbumin promoter (including SDRE, steroid-dependent response element). The Japanese quail ovalbumin promoter was isolated by its high degree of homology to the chicken ovalbumin promoter (GenBank accession number J00895 M24999, base pairs 431-1332). Some deletions were noted in the quail sequence, as compared to the chicken sequence.

Base pairs 4945 - 6092 are a quail ovalbumin signal sequence and ovalbumin gene that corresponds to base pairs

EP 1 592 789 B1

54 - 1201 of GenBank accession number X53964.1. (The STOP codon being omitted).

Base pairs 6093 - 6246 are a TAG sequence containing a gp41 hairpin loop from HIV I an enterokinase cleavage site and a spacer (synthetic).

Base pairs 6247 - 6507 are a proinsulin gene.

5 Base pairs 6514 - 6866 are a synthetic polyadenylation sequence from pGWiz (Gene Therapy Systems) corresponding to base pairs 1920 - 2272 of pGWiz.

Base pairs 6867 - 7303 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 667-235 of pBluescriptII sk(-).

Base pairs 7304- 7379 are the 70 base pairs of the right insertion sequence (IS10) recognized by the transposon Tn10.

10 Base pairs 7380 - 7421 are λ DNA that is residual from pNK2859.

Base pairs 7422 - 8286 are non coding DNA that is residual from pNK2859.

Base pairs 8287 - 10487 are pBlueScript sk(-) base vector (Stratagene, Inc.) corresponding to base pairs 761-2961 of pBluescriptII sk(-).

15 **[0199]** It should be noted that all non-coding DNA sequences described above can be replaced with any other non-coding DNA sequence(s). Missing nucleotide sequences in the above construct represent restriction site remnants.

EXAMPLE 11

20 *Preparation of Transposon-Based Vector pTnMod(Oval/ENT TAG/p146/PA) - Chicken*

[0200] A vector was designed for inserting a p146 gene under the control of a chicken ovalbumin promoter, and a ovalbumin gene including an ovalbumin signal sequence, into the genome of a bird. The vector sequence is provided below as SEQ ID NO:45.

25 Base pairs 1 - 130 are a remainder of F1(-) ori of pBluescriptII sk(-) (Stratagene) corresponding to base pairs 1-130 of pBluescriptII sk(-).

Base pairs 133 - 1777 are a CMV promoter/enhancer taken from vector pGWiz (Gene Therapy Systems) corresponding to base pairs 229-1873 of pGWiz.

30 Base pairs 1780 - 2987 are a transposase, modified from Tn10 (GenBank accession number J01829).

Base pairs 2988-2993 are an engineered stop codon.

Base pairs 2995 - 3410 are a synthetic polyA from pGWiz (Gene therapy Systems) corresponding to base pairs 1922- 2337 of pGWiz.

Base pairs 3415 - 3718 are non coding DNA that is residual from vector pNK2859.

35 Base pairs 3719 - 3761 are λ DNA that is residual from punk2859.

Base pairs 3762 - 3831 are the 70 base pairs of the left insertion sequence (IS10) recognized by the transposon Tn10.

Base pairs 3838 - 4044 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 924-718 of pBluescriptII sk(-).

40 Base pairs 4050 - 4951 are a chicken ovalbumin promoter (including SDRE, steroid-dependent response element) that corresponds to base pairs 431-1332 of the chicken ovalbumin promoter in GenBank Accession Number J00895 M24999.

Base pairs 4958 - 6115 are a chicken ovalbumin signal sequence and Ovalbumin gene that correspond to base pairs 66-1223 of GenBank Accession Number V00383.1 (The STOP codon being omitted).

45 Base pairs 6122 - 6271 are a TAG sequence containing a gp41 hairpin loop from HIV I, an enterokinase cleavage site and a spacer (synthetic).

Base pairs 6272 - 6316 are a p146 sequence (synthetic) with 2 added stop codons.

Base pairs 6324 - 6676 are a synthetic polyadenylation sequence from pGWiz (Gene Therapy Systems) corresponding to base pairs 1920 - 2272 of pGWiz.

50 Base pairs 6682 - 7114 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 667-235 of pBluescriptII sk(-).

Base pairs 7120- 7189 are the 70 base pairs of the right insertion sequence (IS 10) recognized by the transposon Tn10.

Base pairs 7190 - 7231 are λ DNA that is residual from pNK2859.

Base pairs 7232 - 8096 are non coding DNA that is residual from pNK2859.

55 Base pairs 8097 - 10297 are pBlueScript sk(-) base vector (Stratagene, Inc.) corresponding to base pairs 761-2961 of pBluescriptII sk(-).

[0201] It should be noted that all non-coding DNA sequences described above can be replaced with any other non-coding DNA sequence(s). Missing nucleotide sequences in the above construct represent restriction site remnants.

EXAMPLE 12

Preparation of Transposon-Based Vector pTnMod(Oval/ENT TAG/p146/PA) - Quail

5 **[0202]** A vector was designed for inserting a p146 gene under the control of a quail ovalbumin promoter, and a ovalbumin gene including an Ovalbumin signal sequence, into the genome of a bird. The vector sequence is given below as SEQ ID NO:46.

10 Base pairs 1 - 130 are a remainder of F1(-) ori of pBluescriptB sk(-) (Stratagene) corresponding to base pairs 1-130 of pBluescriptII sk(-).
 Base pairs 133 - 1777 are a CMV promoter/enhancer taken from vector pGWiz (Gene Therapy Systems) corresponding to base pairs 229-1873 of pGWiz.
 Base pairs 1780 - 2987 are a transposase, modified from Tn10 (GenBank accession number J01829).
 Base pairs 2988-2993 are an engineered stop codon.
 15 Base pairs 2995 - 3410 are a synthetic polyA from pGWiz (Gene Therapy Systems) corresponding to base pairs 1922-2337 of pGWiz.
 base pairs 3415 - 3718 are non coding DNA that, is residual from vector pUK2859.
 Base pairs 3719 - 3761 are λ DNA that is residual from pNK2859.
 Base pairs 3762 - 3831 are the 70 base pairs of the left insertion sequence (IS10) recognized by the transposon Tn10.
 20 Base pairs 3838 - 4044 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 924-718 of pBluescriptII sk(-).
 Base pairs 4050 - 4938 are the Japanese quail ovalbumin promoter (including SDRE, steroid-dependent response element). The Japanese quail ovalbumin promoter was isolated by its high degree of homology to the chicken ovalbumin promoter (GenBank accession number J00895 M24999, base pairs 431-1332).
 25 Bp 4945 - 6092 are a quail ovalbumin signal sequence and ovalbumin gene that corresponds to base pairs 54 - 1201 of GenBank accession number X53964.1. (The STOP codon being omitted).
 Base pairs 6097 - 6246 are a TAG sequence containing a gp41 hairpin loop from HIV I, an enterokinase cleavage site and a spacer (synthetic).
 Base pairs 6247 - 6291 are a p146 sequence (synthetic) with 2 added stop codons.
 30 Base pairs 6299 - 6651 are a synthetic polyadenylation sequence from pGWiz (Gene Therapy Systems) corresponding to base pairs 1920 - 2272 of pGWiz.
 Base pairs 6657 - 7089 are a multiple cloning site from pBlueScriptII sk(-) corresponding to base pairs 667-235 of pBluescriptII sk(-).
 Base pairs 7095- 7164 are the 70 base pairs of the right insertion sequence (IS10) recognized by the transposon Tn10.
 35 Base pairs 7165 - 7206 are λ DNA that is residual from pNK2859.
 Base pairs 7207 - 8071 are non coding DNA that is residual from pNK2859.
 Base pairs 8072 - 10272 are pBlueScript sk(-) base vector (Stratagene, Inc.) corresponding to base pairs 761-2961 of pBluescriptII sk(-).

40 **[0203]** It should be noted that all non-coding DNA sequences described above can be replaced with any other non-coding DNA sequence(s). Missing nucleotide sequences in the above construct represent restriction site remnants.

EXAMPLE 13

45 *Additional Transposon-Based Vectors for Administration to an Animal*

[0204] The following example provides a description of various transposon-based vectors of the present invention and several constructs that have been made for insertion into the transposon-based vectors of the present invention. These examples are not meant to be limiting in any way. The constructs for insertion into a transposon-based vector
 50 are provided in a cloning vector pTnMCS or pTuMod, both described above.

pTnMCS (CMV-CHOVg-ent-ProInsulin-synPA) (SEQ ID NO: 47)

Bp 1-3670 from vector PTnMCS, bp 1- 3670
 55 Bp 3676 - 5320 CMV promoter/enhancer taken from vector pGWIZ (Gene Therapy Systems), bp 230-1864
 Bp 5327 -6480 Chicken ovalbumin gene taken from GenBank accession # V00383, bp 66-1219
 Bp 6487 - 6636 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
 Bp 6637 - 6897 Human Proinsulin taken from GenBank accession # NM000207, bp 117-377

EP 1 592 789 B1

Bp 6898 - 6942 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and pGWIZ (Gene Therapy Systems)

Bp 6943 - 7295 Synthetic polyA from the cloning vector pGWIZ (Gene Therapy Systems), bp 1920-2271

Bp 7296 -10895 from cloning vector pTnMCS, bp 3716-7315

5

pTnMCS (CMV-prepro-ent-ProInsulin-synPA)

Bp 1- 3670 from vector PTnMCS, bp 1 - 3670

Bp 3676 - 5320 CMV promoter/enhancer taken from vector pGWIZ (Gene Therapy Systems), bp 230-1864

10

Bp 5326 - 5496 Capsite/prepro taken from GenBank accession # X07404, bp 563-733 Bp 5504 - 5652 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site

Bp 5653 - 5913 Human Proinsulin taken from GenBank accession # NM000207, bp 117-377

Bp 5914 - 5958 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and pGWIZ (Gene Therapy Systems)

15

Bp 5959-6310 Synthetic polyA from the cloning vector pGWIZ (Gene Therapy Systems), bp 1920-2271

Bp 6313-9912 from cloning vector pTnMCS, bp 3716-7315

pTnMCS(Chicken OVep+OVg'+ENT+proins+syn polyA)

20

Bp 1-3670 from vector pTnMCS, bp 1- 3670

Bp 3676-4350 Chicken Ovalbumin enhancer taken from GenBank accession #S82527.1 bp 1-675

Bp 4357-5692 Chicken Ovalbumin promoter taken from GenBank accession # J00895M24999 bp 1-1336

Bp 5699-6917 Chicken Ovalbumin gene from GenBank Accession # V00383.1 bp 2-1220. (This sequence includes the 5'UTR, containing putative cap site, bp 5699-5762.)

25

Bp 6924-7073 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site

Bp 7074-7334 Human proinsulin GenBank Accession # NM000207 bp 117-377

Bp 7335-7379 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)

Bp 7380-7731 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920 - 2271

30

Bp 7733-11332 from vector pTnMCS, bp 3716 - 7315

pTnMCS(Chicken OVep+prepro+ENT+proins+syn polyA)

35

Bp 1 - 3670 from cloning vector pTnMCS, bp 1- 3670

Bp 3676 - 4350 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1 bp 1-675

Bp 4357 - 5692 Chicken Ovalbumin promoter taken from GenBank accession # J00895-M24999 bp 1-1336

Bp 5699-5869 Cecropin cap site and prepro, Genbank accession # X07404 bp 563-733

Bp 5876 - 6025 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase, cleavage site

40

Bp 6026 - 6286 Human proinsulin GenBank Accession # NM000207 bp 117-377

Bp 6287 - 6331 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)

Bp 6332 - 6683 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920-2271

Bp 6685 -10284 from cloning vector pTnMCS, bp 3716 - 7315

45

pTnMCS(Quail OVep+QVg'+ENT+proins+syn polyA)

Bp 1- 3670 from cloning vector pTnMCS, bp 1- 3670

50

Bp 3676 - 4333 Quail Ovalbumin enhancer: 658 bp sequence, amplified in-house from quail genomic DNA, roughly equivalent to the far-upstream chicken ovalbumin enhancer, GenBank accession # S82527.1, bp 1-675. (There are multiple base pair substitutions and deletions in the quail sequence, relative to chicken, so the number of bases does not correspond exactly.)

Bp 4340 - 5705 Quail Ovalbumin promoter. 1366 bp sequence, amplified in-house from quail genomic DNA, roughly corresponding to chicken ovalbumin promoter, GenBank accession # J00895-M24999 bp 1-1336. (There are multiple base pair substitutions and deletions between the quail and chicken sequences, so the number of bases does not correspond exactly.)

55

Bp 5712 - 6910 Quail Ovalbumin gene, EMHL accession # X53964, bp 1-1199. (This sequence includes the 5'UTR, containing putative cap site bp 5712-5764.)

EP 1 592 789 B1

Bp 6917- 7066 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
Bp 7067 - 7327 Human proinsulin GenBank Accession # NM000207 bp 117-377
Bp 7328 - 7372 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)
5 Bp 7373 - 7724 Synthetic polyp from the cloning vector gWIZ (Gene Therapy Systems) bp 1920-2271
Bp 7726 -11325 from cloning vector pTnMCS, bp 3716 - 7315

pTnMCS(Quail OVep+prepro+ENT+proins+syn polyA)

10 Bp 1- 3670 from cloning vector pTnMCS, bp 1 - 3670
Bp 3676 -4333 Quail Ovalbumin enhancer: 658 bp sequence, amplified from quail genomic DNA, roughly equivalent to the far- upstream chicken ovalbumin enhancer, GenBank accession #S82527.1, bp 1-675. (There are multiple base pair substitutions and deletions in the quail sequence, relative to chicken, so the number of bases does not correspond exactly.)
15 Bp 4340 - 5705 Quail Ovalbumin promoter. 1366 bp sequence, amplified from quail genomic DNA, roughly corresponding to chicken ovalbumin promoter, GenBank accession # J00895-M24999 bp 1-1336. (There are multiple base pair substitutions and deletions between the quail and chicken sequences, so the number of bases does not correspond exactly.)
Bp 5712-5882 Cecropin cap site and prepro, Genbank accession # X07404 bp 563-733
20 Bp 5889 - 6038 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
Bp 6039 - 6299 Human proinsulin GenBank Accession # NM000207 bp 117-377
Bp 6300 - 6344 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)
Bp 6345 - 6696 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920 - 2271
25 Bp 6698 -10297 from cloning vector pTnMCS, bp 3716 - 7315.

pTnMOD (CMV-prepo-ent-proins-synPA)

Bp 1- 4045 from vector PTnMCS, bp 1-4045
30 Bp 4051 - 5695 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1864
Bp 5701-5871 Capsite/prepro taken from GenBank accession # X07404, bp 563-733
Bp 5879 - 6027 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
Bp 6028-6288 Human Proinsulin taken from GenBank accession # NM000207, bp 117-377
Bp 6289 - 6333 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and
35 pGWIZ (Gene Therapy Systems)
Bp 6334 - 6685 Synthetic polyA from the cloning vector pGWIZ (Gene Therapy Systems), bp 1920-2271
Bp 6687 -10286 from cloning vector pTnMCS, bp 3716-7315

pMnMOD(Chicken OVep+OVg'+ENT+proins+syn polyA)

40 Bp 1 - 4045 from cloning vector pTnMod, bp 1 - 4045
Bp 4051 - 4725 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1 bp 1-675
Bp 4732 - 6067 Chicken Ovalbumin promoter taken from GenBank accession # J00895-M24999 bp 1-1336
Bp 6074 - 7292 Chicken Ovalbumin gene from GenBank Accession # V00383.1 bp 2-1220. (This sequence
45 includes the 5'UTR, containing putative cap site bp 6074-6137.)
Bp 7299 - 7448 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
Bp 7449 - 7709 Human proinsulin GenBank Accession # NM000207 bp 117-377
Bp 7710 - 7754 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)
50 Bp 7755 - 8106 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920-2271
Bp 8108 -11707 from cloning vector pThMod, bp 3716 - 7315

pTnMOD(Chicken OVep+prepro+ENT+proins+syn polyA)

55 Bp 1- 4045 from cloning vector pTnMCS, bp 1- 4045
Bp 4051 - 4725 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1 bp 1-675
Bp 4732 - 6067 Chicken Ovalbumin promoter taken from GenBank accession # J00895-M24999 bp 1-1336
Bp 6074-6244 Cecropin cap site and prepro, Genbank accession # X07404 bp 563-733

EP 1 592 789 B1

Bp 6251 - 6400 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
Bp 6401 - 6661 Human proinsulin GenBank Accession # NM000207 bp 117-377
Bp 6662 - 6706 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)
5 Bp 6707 - 7058 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920 - 2271
Bp 7060 - 10659 from cloning vector pTnMCS, bp 3716 - 7315

pTnMOD(Quail OVep+OVg'+ENT+proins+syn polyA)

10 Bp 1- 4045 from cloning vector pTnMCS, bp 1- 4045
Bp 4051 - 4708 Quail Ovalbumin enhancer: 658 bp sequence, amplified in-house from quail genomic DNA, roughly equivalent to the far-upstream chicken ovalbumin enhancer, GenBank accession # S82527.1, bp 1-675. (There are multiple base pair substitutions and deletions in the quail sequence, relative to chicken, so the number of bases does not correspond exactly.)
15 Bp 4715 - 6080 Quail Ovalbumin promoter: 1366 bp sequence, amplified in-house from quail genomic DNA, roughly corresponding to chicken ovalbumin promoter, GenBank accession # J00995-M24999 bp 1-1336. (There are multiple base pair substitutions and deletions between the quail and chicken sequences, so the number of bases does not correspond exactly.)
Bp 6087 - 7285 Quail Ovalbumin gene, EMBL accession # X53964, bp 1-1199. (This sequence includes the
20 5'UTR, containing putative cap site bp 6087-6139.)
Bp 7292 - 7441 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
Bp 7442 - 7702 Human proinsulin GenBank Accession # NM000207 bp 117-377
Bp 7703 - 7747 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)
25 Bp 7748 - 8099 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920-2271
Bp 8101-11700 from cloning vector pTnMCS, bp 3716-7315

pTnMOD(Quail OVep+prepro+ENT+proins+syn polyA)

30 Bp 1- 4045 from cloning vector pTnMCS, bp 1 - 4045
Bp 4051 - 4708 Quail Ovalbumin enhancer. 658 bp sequence, amplified in-house from quail genomic DNA, roughly equivalent to the far-upstream chicken ovalbumin enhancer, GenBank accession # S82527.1, bp 1-675. (There are multiple base pair substitutions and deletions in the quail sequence, relative to chicken, so the number of bases does not correspond exactly.)
35 Bp 4715 - 6080 Quail Ovalbumin promoter: 1366 bp sequence, amplified in-house from quail genomic DNA, roughly corresponding to chicken ovalbumin promoter, GenBank accession # J00895-M24999 bp 1-1336. (There are multiple base pair substitutions and deletions between the quail and chicken sequences, so the number of bases does not correspond exactly.)
Bp 6087-6257 Cecropin cap site and Prepro, Genbank accession # X07404 bp 563-733
40 Bp 6264 - 6413 Synthetic, spacer sequence and hairpin loop of IDV gp41 with an added enterokinase cleavage site
Bp 6414 - 6674 Human proinsulin GenBank Accession # NM000207 bp 117-377
Bp 6675 - 6719 Spacer DNA, derived as an artifact from the cloning vectors pTOPO Blunt II (Invitrogen) and gWIZ (Gene Therapy Systems)
45 Bp 6720 - 7071 Synthetic polyA from the cloning vector gWIZ (Gene Therapy Systems) bp 1920-2271
Bp 7073 - 10672 from cloning vector pTnMCS, bp 3716 - 7315

pTnMOD (CMV-prepro-ent-hGH-CPA)

50 Bp 1-4045 from vector FTnMOD, bp 1- 4045
Bp 4051-5694 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1873
Bp 5701-5871 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733 Bp 5878-6012 Synthetic spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
Bp 6013-6666 Human growth hormone taken from GenBank accession # V00519, bp 1-654
55 Bp 6673-7080 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058
Bp 7082-10681 from cloning vector pTnMOD, bp 4091-7690

pTnMCS (CHOVep-prepro-ent-hGH-CPA)

EP 1 592 789 B1

Bp 1-3670 from vector PTnMCS, bp 1-3670
Bp 3676-4350 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1, bp 1-675
Bp 4357-5692 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1-1336
5 Bp 5699-5869 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733 Bp 5876-6010 Synthetic
spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
Bp 6011-6664 Human growth hormone taken from GenBank accession # V00519, bp 1-654
Bp 6671-7078 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058
Bp 7080-10679 from cloning vector pTnMCS, bp 3716-7315

10 pTnMCS (CMV-prepro-ent-hGH-CPA)

Bp1- 3670 from vector PTnMCS, bp 1 - 3670
Bp 3676-5319 CMV promoter/enhancer taken from vector pGWIZ (Gene therapy systems), bp 230-1873
Bp 5326-5496 Capsite/Prepro taken from GenBank accession # X07404, bp 563 - 733 Bp 5503-5637 Synthetic
15 spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
Bp 5638-6291 Human growth hormone taken from GenBank accession # V00519, bp 1-654
Bp 6298-6705 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058
Bp 6707-10306 from cloning vector pTnMCS, bp 3716-7315

20 pTnMOD (CHOVep-prepro-ent-hGH-CPA)

Bp 1-4045 from vector PTnMOD, bp 1-4045
Bp 4051-4725 Chicken Ovalbumin enhancer taken from GenBank accession # S82527.1, bp 1-675
Bp 4732-6067 Chicken Ovalbumin promoter taken from GenBank accession # J00899-M24999, bp 1-1336
25 Bp 6074-6244 Capsite/Prepro taken from GenBank accession # X07404, bp 563-733 Bp 6251-6385 Synthetic
spacer sequence and hairpin loop of HIV gp41 with an added enterokinase cleavage site
Bp 6386-7039 Human growth hormone taken from GenBank accession # V00519, bp 1-654
Bp 7046-7453 Conalbumin polyA taken from GenBank accession # Y00407, bp 10651-11058
Bp 7455-11054 from cloning vector pTnMOD, bp 4091-7690

30 PTnMod(CMV/Transposase/ChickOvep/prepro/ProteinA/ConpolyA)

BP 1-130 remainder of F1 (-) on of pBluescriptII sk(-) (Stragagene) bp 1-130.
BP 133-1777 CMV promoter/enhancer taken from vector pGWIZ (Gene Therapy Systems) bp 229-1873.
35 BP 1780-2987 Transposase, modified from Tn10 (GenBank #J01829).
BP 2988-2993 Engineered DOUBLE stop codon.
BP 2994-3343 non coding DNA from vector pNK2859.
BP 3344-3386 Lambda DNA from pNK2859.
BP 3387-3456 70bp of IS10 left from Tn10.
40 BP 3457-3674 multiple cloning site from pBluescriptII sk(-) bp 924-707.
BP 3675-5691 Chicken Ovalbumin enhancer plus promoter from a Topo Clone 10 maxi 040303 (5' XmaI, 3'
BamHI)
BP 5698-5865 prepro with Cap site amplified from cecropin ofpMON200 GenBank # X07404 (5'BamHI, 3'KpnI)
BP 5872-7338 Protein A gene from GenBank# J01786, mature peptide bp 292-1755 (5'KpnI, 3'SacII)
45 BP 7345-7752 ConPolyA from Chicken conalbumin polyA from GenBank # Y00407 bp 10651-11058. (5'SacII,
3'XhoI)
BP 7753-8195 multiple cloning site from pBluescriptII sk(-) bp 677-235.
BP 8196-8265 70 bp of IS10 left from Tn10.
BP 8266-8307 Lambda DNA from pNK2859
50 BP 8308-9151 noncoding DNA from pNK2859
BP 9152-11352 pBluescriptII sk(-) base vector (Stratagene, INC.) bp 761-2961

Appendix A

55 **[0205]**

SEQ ID NO:1 (modified Kozak sequence)
ACCATG

SEQ ID NO:2 (pTnMCS)

1 ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtggttacg cgcagcgtga
 61 cgcgtacact tggcagcgcc ctacgcggcg ctccctttcgc ttctctccct tcctttctcg
 121 ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgatc catatcataa
 181 tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac
 241 tagttattaa tagtaatcaa ttacgggggc attagttcat agcccatata tggagttccg
 301 cgttacataa cttacggtaa atggcccgcg tggctgaccg cccaacgacc ccgcccatt
 361 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca
 421 atgggtggag tatttacggg aaactgccca cttggcagta catcaagtgt atcatatgcc
 481 aagtcacgcc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgccagta
 541 catgacctta tgggactttc ctacttgcca gtacatctac gtattagtca tcgctattac
 601 catggtgatg cgggtttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg
 661 atttccaagt ctccaccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg
 721 ggactttcca aaatgtcgta acaactccgc cccttgacg caaatgggag gtaggcgtgt
 781 acggtgggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg
 841 ccatccacgc tgttttgacc tccatagaag acaccgggac cgtaccagcc tccgcccgcg
 901 ggaacgggtg attggaacgc ggattccccg tgccaagagt gacgtaagta ccgcctatag
 961 actctatagg cacacccctt tggctcttat gcatgtata ctgtttttgg cttggggcct
 1021 atacaccccc gcttccttat gctatagggt atggtatagc ttagcctata ggtgtgggtt
 1081 attgaccatt attgaccact cccctattgg tgacgatact ttccattact aatccataac
 1141 atggctcttt gccacaacta tctctattgg ctatagtcca atactctgtc cttcagagac
 1201 tgacacggac tctgtatttt tacaggatgg ggtccccatt attatttaca aattcacata
 1261 tcaacaacag cgtcccccg tgcgcgcagt ttttataaa catagcgtgg gatctccacg
 1321 cgaatctcgg gtacgtgttc cggacatggg ctcttctcgc gtacggcgcg agcttcaca
 1381 tccgagccct ggtcccatgc ctccagcgcc tcatggtcgc tcggcagctc cttgctccta
 1441 acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag
 1501 gccgtggcgg tagggtatgt gtctgaaaat gagcgtggag attgggctcg cacggtgac
 1561 gcagatggaa gacttaagcg agcggcagaa gaagatgcag gcagctgagt tgttgtatc
 1621 tgataagagt cagaggtaac tcccggtcgc gtgctgttaa cgggtggagg cagtgtagt
 1681 tgagcagtag tctgtgtcgc cgcgcgcgcc accagacata atagctgaca gactaacaga
 1741 ctgttctctt ccatgggtct tttctgcagt caccgtcgga ccatgtgcga actcgatatt
 1801 ttacacgact ctctttacca attctgcccc gaattacact taaaacgact caacagctta
 1861 acgttggctt gccacgcatt acttgactgt aaaactctca ctcttaccga acttggcggt
 1921 aacctgccaa ccaaacggag aacaaaacat aacatcaaac gaatcgaccg attgttaggt
 1981 aatcgtcacc tccacaaaga gcgactcgct gtataccggt ggcatgctag ctttatctgt
 2041 tcgggcaata cgtatcccat tgtacttgtt gactggtctg atattcgtga gcaaaaacga
 2101 cttatgggtat tgcgagcttc agtcgcacta cagcgtcggt ctgttactct ttatgagaaa
 2161 gcggtcccgcc ttccagagca atgttcaaa aaagctcatg accaatttct agccgacctt
 2221 gcgagcattc taccaggtaa caccacaccg ctcatgtca gtgatgctgg ctttaaagt
 2281 ccattgggata aatccgttga gaagctgggt ttgtactgggt taagtgcagt aagaggaaaa
 2341 gtacaatatg cagacctagg agcggaaaaac tggaaaacct tcagcaactt acatgatag
 2401 tcatctagtc actcaaaagac tttaggctat aagaggctga ctaaaagcaa tccaatctca
 2461 tgccaaattc tattgtataa atctcgctct aaaggccgaa aaaatcagcg ctgcacacgg
 2521 actcatgtgc accaccgcgc acctaaaatc tactcagcgt cggcaaaagg gccatgggtt
 2581 ctacgaacta acttacctgt tgaaattcga acacccaaac aacttgttaa tatctattcg
 2641 aagcgaatgc agattgaaga aaccttcga gacttgaaaa gtccctgccta cggactaggc
 2701 ctacgccata gccgaacgag cagctcagag cgttttgata tcatgctgct aatcgccctg
 2761 atgcttcaac taacatggtt gcttgcgggc gttcatgctc agaaaacagg ttgggacaag
 2821 cacttccagg ctaacacagt cagaaatcga aacgtactct caacagttcg cttaggcatg
 2881 gaagttttgc ggcattcttg ctacacaata acaagggaag acttactcgt ggtcgcaacc
 2941 ctactagctc aaaatttatt cacacatggt tacgctttgg ggaaattatg aggggacgcg
 3001 tctagagcga tccgggatct cgggaaaagc gttggtgacc aaaggtgcct tttatcatca
 3061 ctttaaaaaat aaaaaacaat tactcagctg ctgttataag cagcaattaa ttatgattga
 3121 tgccatcacat acaacaaaaa ctgatttaac aaatgggttg tctgccttag aaagtataat
 3181 tgaacattat cttgattata ttattgataa taataaaaac cttatcccta tccaagaagt
 3241 gatgcctatc attggttgga atgaacttga aaaaaattag ccttgaatac attactggta
 3301 aggtaaacgc cattgtcagc aaattgatcc aagagaacca acttaaaagct ttcctgacgg
 3361 aatgttaatt ctggttgacc ctgagcactg atgaatcccc taatgatttt ggtaaaaatc
 3421 attaaagttaa ggtggataca catcttgta tatgatcccc gtaatgtgag ttagctcact
 3481 cattaggcac ccaggcctt acacttatg cttccggctc gtatgttgtg tggaaattgtg

3541 agcggataac aattttcacac aggaaacacagc tatgaccatg attacgccaa ggcgcgaatt
 3601 aaccctcact aaagggaaca aaagctggag ctccaccgcg gtggcggccg ctctagaact
 3661 agtggatccc ccgggctgca ggaattcgat atcaagctta tcgataccgc tgacctcgag
 5 3721 gggggggccc gtacccaatt cgccctatag tgagtctgat taocgcgcgt cactggcgt
 3781 cgttttcaaa cgtcgtgact gggaaaaccc tggcgttacc caacttaate gccttgacgc
 3841 acatccccct ttccgagctt ggcgtaatag cgaagaggcc cgcaccgacg gcccttccca
 3901 acagttgcgc agcctgaatg gcgaatggaa attgtaagcg ttaatatatt gttaaaattc
 3961 gcgttaaat tttgttaaat cagctcattt ttaaccaat aggcggaat cggaataatc
 4021 ccttataaat caaaagaata gaccgagata ggggtgagtg ttgttccagt ttggaacaag
 4081 agtccactat taaagaacgt ggactccaac gtcaaaaggc gaaaaacgt ctatcagggc
 10 4141 gatggcccac tactccggga tcatatgaca agatgtgat ccaccttaac ttaatgattt
 4201 ttaccaaaat cattagggga ttcacagtg ctcagggtca acgagaatta acattccgtc
 4261 aggaagcctt atgatgatga tgtgcttaaa aacttactca atggctgggt atgcatacgc
 4321 caatacatgc gaaaaaccta aaagagcttg ccgataaaaa aggccaatat attgctattt
 4381 acgcggcgtt tttattgagc ttgaaagata aataaaatag ataggtttta tttgagacta
 15 4441 aatcttcttt atcgtaaaaa atgccccttt gggttatcaa gagggtcatt atatttcgcg
 4501 gaataacatc atttgggtgac gaaataacta agcacttgct tcctgtttac tcccctgagc
 4561 ttgagggggt aacatgaagg tcatcgatag caggataata atacagtaaa acgctaaacc
 4621 aataatccaa atccagccat cccaaattgg tagtgaatga ttataaataa cagcaaacag
 4681 taatgggcca ataacacccg ttgcatgggt aaggctcacc aataatccct gtaagcacc
 4741 ttgctgatga ctctttgttt gगतगacat cactccctgt aatgcaggta aagcgatccc
 4801 accaccagcc aataaaatta aaacagggaa aactaaccaa ccttcagata taaacgctaa
 20 4861 aaaggccaaat gcactactat ctgcaataaa tccgagcagt actgcggtt tttcgccat
 4921 ttagtggcta ttcttctgca cacaaggct tggaaactg agtgtaaaag accaagaccc
 4981 gtaataaaaa gcccaaccat atgctattca tcatcacgat ttctgtataa gcaccacacc
 5041 gtgctggatt ggctatcatt gcgctgaaat aataatcaac aaatggcatc gttaaataag
 5101 tgatgtatac cgatcagctt ttgttccctt tagtgagggt taattgcgcg cttggcgtaa
 25 5161 tcatgggtcat agctgtttcc tgtgtgaaat tgttatccgc tcacaattcc acacaacata
 5221 cgagccggaa gcataaagt taaagcctgg ggtgcctaat gagtgagcta actcacatta
 5281 attgcgttgc gctcactgct cgctttccag tccggaaacc tgcgtgcca gctgcattaa
 5341 tgaatcggcc aaocgcggg gagaggcggt ttgcgtattg ggcgtctctc cgttctctc
 5401 ctactgact cgtcgcgtc ggtcgttcgg ctgcggcgag cgggtatcagc tcatcacaag
 5461 gcggtaatac ggttatccac agaatacagg gataacgcag gaaagaacat gtgagcaaaa
 5521 ggcacgcaaa aggcacaggaa ccgtaaaaag cgcgcgttgc tggcggtttt ccataggctc
 30 5581 cgcctccctg acgagcatca caaaaatcga cgtcgaatc agaggtggcg aaaccgcaca
 5641 ggactataaa gataccaggc gtttccccct ggaagctccc tgcgtgcgct tccgttccg
 5701 accctgcccg ttaccggata cctgtccgct tttctccctt cgggaagcgt ggcgcttct
 5761 ctactgcac gctgtaggta tctcagttcg gttaggtcg ttccgtccaa gctgggctgt
 5821 gtgcacgaac ccccggttca gcccgaccgc tgcgccttat cgggtaacta tgccttgag
 5881 tccaaccogg taagacagca ctatcgcca ctggcagcag ccactggtta caggattagc
 35 5941 agagcagggt atgtaggcgg tgctacagag ttcttgaaat ggtggcctaa ctacggctac
 6001 actagaaggga cagtatttgg tatctgcgct ctgctgaagc cagttacctt cggaaaaaga
 6061 gttggtagct cttgatccgg caaacaacc accgctggta gcggtggttt tttgtttgc
 6121 aagcagcaga ttacgcgcag aaaaaaaggga tctcaagaag atcctttgat cttttctacg
 6181 gggctctgacg ctacgtggaa cgaaaactca cgttaaggga ttttggctcat gagattatca
 6241 aaaaaggatct tcacctagat ccttttaaat taaaaatgaa gttttaaatc aatctaaagt
 40 6301 atatatgagt aaacttggct tgacagttac caatgcttaa tcagtgggc accatatctca
 6361 gcgatctgct tatttcgctt atccatagtt gctgactcc cgtcgtgta gataactacg
 6421 atacgggagg gcttaccatc tggccccagt gctgcaatga taccgcgaga cccagctca
 6481 ccggtccag atttatcagc aataaacccag ccagccggaa gggccgagcg cagaagtgg
 6541 cctgcaactt tatccgcctc catccagttc attaatgtt gccgggaagc tagagtaagt
 6601 agtgcggcag ttaatagttt gcgcaacgtt gttgccattg ctacaggcat cgtgggtgca
 6661 cgtcgtcgt ttggtatggc ttcattcagc tccggttccc aacgatcaag gcgagttaca
 45 6721 tgatccccc tgttgtgcaa aaaagcgggt agctccttcg gtcctccgat cgttgcaga
 6781 agtaagttgg ccgcagttt atcactcatg gttatggcag cactgcataa ttctcttact
 6841 gtcatgccat ccgtaagatg cttttctgtg actgggtgag actcaacca gtcattctga
 6901 gaatagtgta tgcggcgacc gagttgctct tgcgcggcgt caatacggga taataccgcg
 6961 ccacatagca gaactttaaa agtgctcatc attggaaaac gttcttcggg gcgaaaactc
 7021 tcaaggatct taccgctggt gagatccagt tcgatgtaac ccactcgtgc acccaactga
 50 7081 tcttcagcat cttttacttt caccagcgtt tctgggtgag caaaaacagg aaggcaaat
 7141 gccgcaaaaa agggaaataag ggcgacacgg aaatgttgaa tactcatact cttcttttt
 7201 caatattatt gaagcattta tcagggttat tgtctcatga gcggatacat atttgaatgt
 7261 atttagaaaa ataaacaaat aggggttccg cgcacatttc cccgaaaagt gccac

SEQ ID NO:3 (pTnMod)

	CTGACGCGCC	CTGTAGCGGC	GCATTAAGCG	CGGCGGGTGT	GGTGGTTACG	50
	CGCAGCGTGA	CCGCTACACT	TGCCAGCGCC	CTAGCGCCCG	CTCCTTTCGC	100
	TTTCTTCCCT	TCCTTTCTCG	CCACGTTGCG	CGGCATCAGA	TTGGCTATTG	150
5	GCCATTGCGT	ACGTTGTATC	CATATCATAA	TATGTACATT	TATATTGGCT	200
	CATGTCCAAC	ATTACGCGCA	TGTTGACATT	GATTATTGAC	TAGTTATTAA	250
	TAGTAATCAA	TTACGGGGTC	ATTAGTTCAT	AGCCCATATA	TGGAGTTCG	300
	CGTTACATAA	CTTACGGTAA	ATGGCCCGCC	TGGCTGACCG	CCCAACGACC	350
	CCCGCCCAT	GACGTCAATA	ATGACGTATG	TTCCCATAGT	AACGCCAATA	400
	GGGACTTTCC	ATTGACGTCA	ATGGGTGGAG	TATTTACGGT	AAACTGCCCA	450
10	CTTGGCAGTA	CATCAAGTGT	ATCATATGCC	AAGTACGCCC	CCTATTGACG	500
	TCAATGACGG	TAAATGGCCC	GCCTGGCATT	ATGCCAGTA	CATGACCTTA	550
	TGGGACTTTC	CTACTTGGCA	GTACATCTAC	GTATTAGTCA	TCGCTATTAC	600
	CATGGTGATG	CGGTTTGGC	AGTACATCAA	TGGCGGTGGA	TAGCGGTTTG	650
	ACTCACGGGG	ATTTCCAAGT	CTCCACCCCA	TTGACGTCAA	TGGGAGTTTG	700
15	TTTTGGCACC	AAAATCAACG	GGACTTTCCA	AAATGTCGTA	ACAACTCCGC	750
	CCATTGACG	CAAATGGGCG	GTAGGCGTGT	ACGGTGGGAG	GTCTATATAA	800
	GCAGAGCTCG	TTTAGTGAAC	CGTCAGATCG	CCTGGAGACG	CCATCCACGC	850
	TGTTTTGACC	TCCATAGAAG	ACACCGGGAC	CGATCCAGCC	TCCGCGGCCG	900
	GGAAOOGTGC	ATTGGAACGC	GGATTCCCG	TGCCAAGAGT	GACGTAAGTA	950
	CCGCCTATAG	ACTCTATAGG	CACACCCCTT	TGGCTCTTAT	GCATGCTATA	1000
20	CTGTTTTTGG	CTTGGGGCCT	ATACACCCCC	GCITCCTTAT	GCTATAGGTG	1050
	ATCGTATAGC	TTAGCCTATA	GGTGTGGGTT	ATTGACCATT	ATTGACCACT	1100
	CCCTATTGG	TGACGATACT	TTCCATTACT	AATCCATAAC	ATGGCTCTTT	1150
	GCACAACATA	TCTCTATTGG	CTATATGCCA	ATACTCTGTC	CTTCAGAGAC	1200
	TGACACGGAC	TCTGTATTTT	TACAGGATGG	GGTCCCATTT	ATTATTTACA	1250
25	AATTCACATA	TACAACAACG	CCGTCCCCCG	TGCCCGCAGT	TTTTATTAAA	1300
	CATAGCGTGG	GATCTCCACG	CGAATCTCGG	GTACGTGTTC	CGGACATGGG	1350
	CTCTTCTCCG	GTAGCGGCGG	AGCTTCCACA	TCCGAGCCCT	GGTCCCATGC	1400
	TCCAGCGGC	TCATGGTCCG	TCCGCGAGTC	CTTGCTCCTA	ACAGTGGAGG	1450
	CCAGACTTAG	GCACAGCACA	ATGCCACCA	CCACCAGTGT	GCCGCACAAG	1500
	GCCGTGGCGG	TAGGGTATGT	GTCTGAAAAT	GAGCGTGGAG	ATTGGGCTCG	1550
30	CACGGCTGAC	GCAGATGGAA	GACTTAAGGC	AGCGCGAGAA	GAAGATGCAG	1600
	GCAGCTGAGT	TGTTGTATTC	TGATAAGAGT	CAGAGGTAAC	TCCCGTTGCG	1650
	GTGCTGTIAA	CGGTGGAGGG	CAGTGTAGTC	TGAGCAGTAC	TCGTTGCTGC	1700
	CGCGCGCGCC	ACCAGACATA	ATAGCTGACA	GACTAACAGA	CTGTTCTCTT	1750
	CCATGGGTCT	TTTCTGCAGT	CACCGTCGGA	CCATGTGTGA	ACTTGATATT	1800
35	TTACATGATT	CTCTTTACCA	ATTCTGCCCC	GAATTACACT	TAAAACGACT	1850
	CAACAGCTTA	ACGTTGGCTT	GCCACGCATT	ACTTGACTGT	AAAACCTCTCA	1900
	CTCTTACCBA	ACTTGGCCGT	AACCTGCCAA	CCAAAGCGAG	AACAAAACAT	1950
	AAATCAAAAC	GAATCGACGC	ATTGTTAGGT	AATCGTCAAC	TCCACAAAGA	2000
	GCGACTCGCT	GTATACCGTT	GGCATGCTAG	CTTTATCTGT	TCCGGAATAC	2050
	GATGCCCAT	GTACTTGTG	ACTGGTCTGA	TATTCGTGAG	CAAAAACGAC	2100
40	TTATGGTATT	GCGAGCTTCA	GTCGCACTAC	ACGGTCGTTC	TGTTACTCTT	2150
	TATGAGAAAG	CGTTCCCGCT	TTCAGAGCAA	TGTTCAAAGA	AAGCTCATGA	2200
	CCAATTTCTA	GCCGACCTTG	CGAGCATTCT	ACCGAGTAAC	ACCACACCGC	2250
	TCATTGTGAG	TGATGCTGGC	TTTAAAGTGC	CATGGTATAA	ATCCGTTGAG	2300
	AAGCTGGGTT	GGTACTGGTT	AAGTCGAGTA	AGAGGAAAAG	TACAATATGC	2350
45	AGACCTAGGA	GCGGAAAAC	GGAAACCTAT	CAGCAACTTA	CATGATATGT	2400
	CATCTAGTCA	CTCAAAGACT	TTAGGCTATA	AGAGGCTGAC	TAAAAGCAAT	2450
	CCAATCTCAT	GCCAAATTCT	ATTGTATAAA	TCTCGCTCTA	AAGGCCGAAA	2500
	AAATCAGCGC	TCGACACGGA	CTCATTGTCA	CCACCCGTCA	CCTAAAATCT	2550
	ACTCAGCGTC	GGCAAAGGAG	CCATGGGTTC	TAGCAACTAA	CTTACCTGTT	2600
	GAAATTCGAA	CACCCAAACA	ACTTGTAAAT	ATCTATTGCA	AGCGAATGCA	2650
50	GATTGAAGAA	ACCTTCCGAG	ACTTGAAAAG	TCCTGCCTAC	GGACTAGGCC	2700
	TACGCCATAG	CCGAACGAGC	AGCTCAGAGC	GTTTTGATAT	CATGCTGCTA	2750
	ATCGCCCTGA	TGCTTCAACT	AACATGTTGG	CTTGCGGGCG	TTCATGCTCA	2800
	GAAACAAGGT	TGGGACAAGT	ACTTCCAGGC	TAAACAGTC	AGAAATCGAA	2850
	ACGTAATCTC	AACAGTTCGC	TTAGGCATGG	AAGTTTTGCG	GCATTCTGGC	2900
	TACACAATAA	CAAGGGAAGA	CTTACTCGTG	GCTGCAACCC	TACTAGCTCA	2950
55	AAATTTATTC	ACACATGGTT	ACGCTTTGGG	GAAATTATGA	TAATGATCCA	3000
	GATCACTTCT	GGCTAATAAA	AGATCAGAGC	TCTAGAGATC	TGTGTGTTGG	3050

	TTTTTGTGG	ATCTGCTGTG	CCTTCTAGTT	GCCAGCCATC	TGTTGTTTGC	3100
	CCCTCCCCCG	TGCCCTTCCTT	GACCCCTGGAA	GGTGCCACTC	CCACTGTCCT	3150
	TTCCTAATAA	AATGAGGAAA	TTGCATCGCA	TTGTCTGAGT	AGGTGTCATT	3200
5	CTATTCTGGG	GGGTGGGGTG	GGGCAGCACA	GCAAGGGGGA	GGATTGGGAA	3250
	GACAATAGCA	GGCATGCTGG	GGATGCGGTG	GGCTCTATGG	GTACCTCTCT	3300
	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CTCTCGGTAC	CTCTCTCTCT	3350
	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CGGTACCAGG	TGCTGAAGAA	3400
	TTGACCCGGT	GACCAAAGGT	GCCTTTTATC	ATCACTTTAA	AAATAAAAAA	3450
	CAATTACTCA	GTGCCTGTTA	TAAGCAGCAA	TTAATTATGA	TTGATGCCTA	3500
10	CATCACAACA	AAAACCTGATT	TAACAAATGG	TTGGTCTGCC	TTAGAAAAGTA	3550
	TATTTGAACA	TTATCTTGAT	TATATTATTG	ATAATAATAA	AAACCTTATC	3600
	CCTATCCAAG	AAGTGATGCC	TATCATTGGT	TGGAATGAAC	TTGAAAAAAA	3650
	TTAGCCTTGA	ATACATTACT	GGTAAGGTAA	ACGCCATTGT	CAGCAAATTG	3700
	ATCCAAGAGA	ACCAACTTAA	AGCTTTCCTG	ACGGAATGTT	AATTCTCGTT	3750
	GACCCGTAGC	ACTGATGAAT	CCCCTAATGA	TTTGGTAAA	AATCATTAAAG	3800
15	TTAAGGTGGA	TACACATCTT	GTCAATATGAT	CCCGGTAATG	TGAGTTAGCT	3850
	CACTCATTAG	GCACCCAGG	CTTTACACTT	TATGCTTCCG	GCTCGTATGT	3900
	TGTGTGGAAT	TGTGAGCGGA	TAACAATTTC	ACACAGGAAA	CAGCTATGAC	3950
	CATGATTACG	CCAAGCGCGC	AATTAACCCCT	CACTAAAGGG	AACAAAAGCT	4000
	GGAGCTCCAC	CGCGGTGGCG	GCCGCTCTAG	AACTAGTGGA	TCCCCGGGGC	4050
20	TGCAGGAATT	CGATATCAAG	CTTATCGATA	CCGCTGACCT	CGAGGGGGGG	4100
	CCCGGTACCC	AATTCGCCCT	ATAGTGAGTC	GTATTACGGC	CGCTCACTGG	4150
	CCGTCGTTTT	ACAACGTCGT	GACTGGGAAA	ACCTGGCGT	TACCCAACCT	4200
	AATCGCCTTG	CAGCACATCC	CCCTTTCGCC	AGCTGGCGTA	ATAGCGAAGA	4250
	GGCCCGCACC	GATCGCCCTT	CCCAACAGTT	GCGCAGCCTG	AATGGCGAAT	4300
	GGAAATTGTA	AGCGTTAATA	TTTTGTTAAA	ATTGCGGTAA	AATTTTGTG	4350
25	AAATCAGCTC	ATTTTFTAAC	CAATAGGCCG	AAATCGGCAA	AATCCCTTAT	4400
	AAATCAAAAG	AATAGACCGA	GATAGGGTGT	AGTGTGTGTC	CAGTTTGGAA	4450
	CAAGAGTCCA	CTATTAAAGA	ACGTGGACTC	CAACGTCAAA	GGGCGAAAAA	4500
	CCGTCTATCA	GGGCGATGGC	CCACTACTCC	GGGATCATAT	GACAAGATGT	4550
	GTATCCACCT	TAACCTAATG	ATTTTACCA	AAATCATTAG	GGGATTCATC	4600
30	AGTGCTCAGG	GTCAACGAGA	ATTAACATTC	CGTCAGGAAA	GCTTATGATG	4650
	ATGATGTGCT	TAAAACTTA	CTCAATGGCT	GGTTATGCAT	ATCGCAATAC	4700
	ATGCGAAAAA	CCTAAAAGAG	CTTGCCGATA	AAAAAGGCCA	ATTTATGTCT	4750
	ATTTACGCGG	GCTTTTATT	GAGCTTGAAA	GATAAATAAA	ATAGATAGGT	4800
	TTTATTTGAA	GCTAAATCTT	CTTTATCGTA	AAAAATGCC	TCTTGGGTTA	4850
	TCAAGAGGGT	CATTATATTT	CGCGGAATAA	CATCATTTGG	TGACGAAATA	4900
35	ACTAAGCACT	TGTCTCCTGT	TTACTCCCCT	GAGCTTGAGG	GGTTAACATG	4950
	AAGGTCATCG	ATAGCAGGAT	AATAATACAG	TAAACGCTA	AAACCAATAA	5000
	CCAAATCCAG	CCATCCCAAA	TTGGTAGTGA	ATGATTATAA	ATAACAGCAA	5050
	ACAGTAATGG	GCCAATAACA	CCGGTTGCAT	TGGTAAGGCT	CACCAATAAT	5100
	CCCTGTAAAG	CACCTTGCTG	ATGACTCTTT	GTTTGGATAG	ACATCACTCC	5150
40	CTGTAATGCA	GGTAAAGCGA	TCCCACCACC	AGCCAATAAA	ATTAAACAG	5200
	GGAAAACTAA	CCAACCTTCA	GATATAAAGC	CTAAAAAGGC	AAATGCACTA	5250
	CTATCTGCAA	TAAATCCGAG	CAGTACTGCC	GTTTTCCTGC	CCATTTAGTG	5300
	GCTATTCTTC	CTGCCACAAA	GGCTTGGAAT	ACTGAGTGTA	AAAGACCAAG	5350
	ACCCGTAAATG	AAAAGCCAAC	CATCATGCTA	TTTCATCATCA	CGATTTCTGT	5400
	AATAGCACCA	CACCGTGCTG	GATTGGCTAT	CAATGCGCTG	AAATAATAAT	5450
45	CAACAAATGG	CATCGTTAAA	TAAGTGATGT	ATACCGATCA	GCTTTTGTTC	5500
	CCTTTAGTGA	GGGTTAATTG	CGCGCTTGGC	GTAATCATGG	TCATAGCTGT	5550
	TTCTGTGTG	AAATTGTTAT	CCGCTCACAA	TTCCACACAA	CATACGAGCC	5600
	GGAAGCATAA	AGTGTAAGAG	CTGGGGTGCC	TAATGAGTGA	GCTAACTCAC	5650
	ATTAATTGCG	TGCGCTCAC	TGCCCGCTTT	CCAGTCGGGA	AACCTGTGCT	5700
50	GCCAGCTGCA	TTAATGAATC	GGCCAACGCG	CGGGGAGAGG	CGGTTTGGCT	5750
	ATTGGGCGCT	CTTCCGCTTC	CTCGCTCACT	GACTCGCTGC	GCTCGGTCGT	5800
	TCGGCTGCGG	CGAGCGGTAT	CAGCTCACTC	AAAGGCGGTA	ATACGGTTAT	5850
	CCACAGAATC	AGGGGATAAC	GCAGGAAAGA	ACATGTGAGC	AAAAGGCCAG	5900
	CAAAAGGCCA	GGAACCGTAA	AAAGGCCGCG	TTGCTGGCGT	TTTTCCATAG	5950
	GCTCCGCCCC	CCTGACGAGC	ATCACAATAA	TCGACGCTCA	AGTCAGAGGT	6000
55	GGCGAAACCC	GACAGGACTA	TAAAGATACC	AGGCGTTTCC	CCCTGGAAGC	6050
	TCCCTCGTGC	GCTCTCCTGT	TCCGACCCTG	CCGCTTACCG	GATACCTGTG	6100

5 CGCCTTTCTC CCTTCGGGAA GCGTGGCGCT TTCTCATAGC TCACGCTGTA 6150
 GGTATCTCAG TTCGGTGTAG GTCGTTCGCT CCAAGCTGGG CTGTGTGCAC 6200
 GAACCCCCCG TTCAGCCCGA CCGCTGCGCC TTATCCGGTA ACTATCGTCT 6250
 TGAGTCCAAC CCGGTAAGAC ACGACTTATC GCCACTGGCA GCAGCCACTG 6300
 GTAACAGGAT TAGCAGAGCG AGGTATGTAG GCGGTGCTAC AGAGTTCTTG 6350
 AAGTGGTGGC CTAACACGG CTACACTAGA AGGACAGTAT TTGGTATCTG 6400
 CGCTCTGCTG AAGCCAGTTA CCTTCGGAAA AAGAGTTGGT AGCTCTTGAT 6450
 CCGGCAACA AACCAACGCT GGTAGCGGTG GTTTTTTGT TTGCAAGCAG 6500
 CAGATTACGC GCAGAAAAA AGGATCTCAA GAAGATCCTT TGATCTTTTC 6550
 10 TACGGGGTCT GACGCTCAGT GGAACGAAAA CTCACGTAA GGGATTTTGG 6600
 TCATGAGATT ATCAAAAAGG ATCTTCACCT AGATCCTTTT AAATTAATAA 6650
 TGAAGTTTAA AATCAATCTA AAGTATATAT GAGTAACTT GGTCTGACAG 6700
 TTACCAATGC TTAATCAGTG AGGCACCTAT CTCAGCGATC TGTCTATTTT 6750
 GTTCATCCAT AGTTGCCTGA CTCCCGTCTG TGTAGATAAC TACGATACGG 6800
 GAGGGCTTAC CATCTGGCCC CAGTGTGCA ATGATACCGC GAGACCCACG 6850
 15 CTCACCGGCT CCAGATTTAT CAGCAATAAA CCAGCCAGCC GGAGGGGCGG 6900
 AGCGCAGAAG TGGTCCTGCA ACTTTATCCG CCTCCATCCA GTCTATTAAT 6950
 TGTTCGCCGG AAGCTAGAGT AAGTAGTTCG CCAGTTAATA GTTTGCGCAA 7000
 CGTTGTTGCC ATTGCTACAG GCATCGTGGT GTCACGCTCG TCGTTTGGTA 7050
 TGGCTTCATT CAGCTCCGGT TCCCAACGAT CAAGGCGAGT TACATGATCC 7100
 20 CCCATGTTGT GCAAAAAAGC GGTTAGCTCC TTCGGTCTC CGATCGTTGT 7150
 CAGAAGTAAG TTGGCCGCG TGTATCACT CATGGTTATG GCAGCACTGC 7200
 ATAATTCTCT TACTGTCTAG CCATCCGTAA GATGCTTTTC TGTGACTGGT 7250
 GAGTACTCAA CCAAGTCATT CTGAGAATAG TGTATGCGGC GACCGAGTTG 7300
 CTCTTGCCCG GCGTCAATAC GGGATAATAC CGCGCCACAT AGCAGAACTT 7350
 TAAAAGTGCT CATCATTGGA AAACGTTCTT CGGGGCGAAA ACTCTCAAGG 7400
 25 ATCTTACCGC TGTGAGATC CAGTTCGATG TAACCCACTC GTGCACCCAA 7450
 CTGATCTTCA GCATCTTTTA CTTTCACCAG CGTTTCTGGG TGAGCAAAAA 7500
 CAGGAAGGCA AAATGCCGCA AAAAAGGGAA TAAGGGCGAC ACGGAAATGT 7550
 TGAATACTCA TACTCTTCTT TTTTCAATAT TATTGAAGCA TTTATCAGGG 7600
 TTATTGTCTC ATGAGCGGAT ACATATTTGA ATGTATTTAG AAAAATAAAC 7650
 30 AAATAGGGGT TCCGCGCACA TTTCCCGGAA AAGTGCCAC 7689

SEQ ID NO:4 (a Kozak sequence)
ACCATGG

35 SEQ ID NO: 5 (a Kozak sequence)
ACCATGT

SEQ ID NO:6 (a Kozak sequence)
AAGATGT

40 SEX ID NO:7 (a Kozak sequence)
ACGATGA

SEQ ID NO:8 (a Kozak sequence)
AAGATGG

SEQ ID NO:9 (a Kozak sequence)
GACATGA

50 SEQ ID NO:10 (a Kozak sequence)
ACCATGA

55 **SEQ ID NO:11 (conalbumin polyA)**
 tctgccattg ctgcttcctc tgccttcct cgtcactctg aatgtggctt cttegetact
 gccacagcaa gaaataaaat ctcaacatct aaatgggttt cctgagggtt ttcaagagtc

5

10

15

20

25

30

35

40

45

50

55

SEQ ID NO:14
(vitellogenin promoter)

5 TGAATGTGTT CTTGTGTTAT
CAATATAAAT CACAGTTAGT
GATGAAGTTG GCTGCAAGCC
TGCATCAGTT CAGCTACTTG
GCTGCATTTT GTATTGTT
10 CTGTAGGAAA TGCAAAAGGT
TCTAGGCTGA CCTGCACTTC
TATCCCTCTT GCCTTACTGC
TGAGAATCTC TGCAGGTTTT
AATTGTTTAC ATTTTGCTCC
CATTTACTTT GGAAGATAAA
15 ATATTTACAG AATGCTTATG
AAACCTTTGT TCATTTAAAA
ATATTCCTGG TCAGCGTGAC
CGGAGCTGAA AGAACACATT
GATCCCGTGA TTTCAATAAA
TACATATGTT CCATATATTG
20 TTTCTCAGTA GCCTCTTAAA
TCATGTGCGT TGGTGACAT
ATGAATACAT GAATAGCAAA
GGTTTATCTG GATTACGCTC
TGGCCTGCAG GAATGGCCAT
25 AAACCAAGC TGAGGGAAGA

30 GGGAGAGTAT AGTCAATGTA
GATTATACTG ATTGCTGATT
GGGTTATTAT CAGCTAGATA
ACAACTTGGG TCAGGTGCCA
GGTCAACATA ACCTGGGCAA
AACCAGTCTC ATCTGTGGCA
GGACCATGTA CCAGCAGCCA
35 GCCGTGACCC AATCTAGGAA
AGCAAGTAGC ACATCAATTT
TAAATTTATT GTAAATGCCG
TAGTAGAAGT GTTTTACTGT
GATACATTGA AACTTCTGGT
CAATCAGAAA AAGGTTTTTT
40 ATCAGAGATG CCAAGGTATT
ATTTGATTTT CTTTATTGCG
CGTGAAGAGA ATTTATGATT
GCAAAAAGAG GAGTGTTTAC
ATAAACTGAT AAAAACTTG
AGGAATTCAG CAGAAAACAG
45 CCACGTGTTC CTGAACATT
TTCCATAAAA GTCTCACCAT
GCCTGGCAGA GCCCTATTCA
CCTTCGCT

50

55

SEQ ID NO:15 (fragment of ovalbumin promoter - chicken)

GAGGTCAGAAT GGTTCCTTTA CTGTTTGTC AATCTATTAT TTCAATACAG
 AACAAATAGCT TCTATAACTG AAATATATTT GCTATTGTAT ATTATGATTG
 TCCTCGAAC CATGAACACT CCTCCAGCTG AATTTACAA TTCCTGTGTC
 ATCTGCCAGG CCATTAAGTT ATTCATGGAA GATCTTTGAG GAACACTGCA
 AGTTCATATC ATAAACACAT TTGAAATGA GTATTGTTTT GCATTGTATG
 GAGCTATGTT TTGCTGTATC CTCAGAAAAA AAGTTTGTTA TAAAGCATTG
 ACACCCATAA AAAGATAGAT TTAAATATTC CAGCTATAGG AAAGAAAGTG
 CGTCTGCTCT TCACTCTAGT CTCAGTTGGC TCCTTCACAT GCATGCTTCT
 TTATTTCTCC TATTTTGTC AAAAAATAAT AGGTCACGTC TTGTTCTCAC
 TTATGTCTG CCTAGCATGG CTCAGATGCA CGTTGTAGAT ACAAGAAGGA
 TCAAATGAAA CAGACTTCTG GTCTGTACT ACAACCATAG TAATAAGCAC
 ACTAACTAAT AATTGCTAAT TATGTTTTCC ATCTCTAAGG TTCCACATT
 TTCTGTTTT CTTAAAGATC CCATTATCTG GTGTAACTG AAGCTCAATG
 GAACATGAGC AATATTTCCC AGTCTTCTCT CCCATCRAAC AGTCTGATG
 GATTAGCAGA ACAGGCAGAA AACACATTGT TACCAGAAT TAAAACTAA
 TATTGCTCT CATTCAATC CAAAATGGAC CTATTGAAAC TAAATCTAA
 CCCAATCCCA TTAAATGATT TCTATGGCGT CAAAGGTCAA ACTTCTGAAG
 GGAACCTGTG GGTGGGTCAC AATTCAGGCT ATATATTCCC CAGGGCTCAG

SEQ ID NO:16 (chicken ovalbumin enhancer)

ccgggctgca gaaaaatgcc aggtggacta tgaactcaca tccaaaggag
 cttgacctga tacctgattt tcttcaaact ggggaaacaa cacaatccca caaaacagct
 cagagagaaa ccatactga tggctacagc accaaggat gcaatggcaa tccattcgac
 attcatctgt gacctgagca aaatgattta tctctccatg aatgggtgct tctttccctc
 atgaaaaggc aatttccaca ctcaaatat gcaacaaaga caaacagaga acaattaatg
 tgctccttcc taatgtcaaa attgtagtgg caaagaggag aacaaaatct caagttctga
 gtaggtttta gtgattggat aagaggcttt gacctgtgag ctcacctgga cttcatatcc
 ttttgataa aaagtgttt tataactttc aggtctccga gtctttatcc atgagactgt
 tgggttaggg acagacccac aatgaaatgc ctggcatagg aaagggcagc agagccttag
 ctgacctttt cttgggacaa gcattgtcaa acaatgtgtg acaaaactat ttgtactgct
 ttgcacagct gtgctgggca gggcaatcca ttgccaccta tcccaggtaa ccttccaact
 gcaagaagat tgttgcttac tctctctaga

SEQ ID NO:17 (5' untranslated region)

GTGGATCAACATACAGCTAGAAAGCTGTATTGCCTTTAGCACTCAAGCTCAAAGACAACCTCAGAGTTC
 ACC

SEQ ID NO:18 (putative cap site)

ACATACAGCTAG AAAGCTGTAT TGCCTTTAGC ACTCAAGCTC AAAAGACAAC TCAGAGrrcA

SEQ ID NO:19 (Chicken Ovalbumin Signal Sequence)
 ATG GGCTCCATCG GCGCAGCAAG CATGGAATTT TGTTTGATG TATTCAAGGA GCTCAAAGTC
 CACCATGCCA ATGAGAACAT CTTCTACTGC CCCATTGCCA TCATGTCAGC TCTAGCCATG
 GTATACCTGG GTGCAAAAGA CAGCACCAGG ACACAGATAA ATAAGGTTGT TCGCTTTGAT
 AAACCTTCAG GATTGCGAGA CAGTATTGAA GCTCAGTGTG GCACATCTGT AAACGTTTAC
 TCTTCACCTA GAGACATCCT CAACCAATC ACCAAACCAA ATGATGTTTA TTCGTTTCAGC
 CTTGCCAGTA GACTTTATGC TGAAGAGAGA TACCCAATCC TGCCAGAATA CTTGCAGTGT
 GTGAAGGAAC TGTATAGAGG AGGCTTGGA CCTATCAACT TTCAAACAGC TGCAGATCAA
 GCCAGAGAGC TCATCAATTC CTGGGTAGAA AGTCAGACAA ATGGAATTAT CAGAAATGTC
 CTTCAGCCAA GCTCCGTGGA TTCTCAAAC GCAATGGTTC TGGTTAATGC CATTGTCTTC
 AAAGGACTGT GGGAGAAAAC ATTAAAGGAT GAAGACACAC AAGCAATGCC TTTCAGAGTG
 ACTGAGCAAG AAAGCAAACC TGTGCAGATG ATGTACCAGA TTGGTTTATT TAGAGTGGCA
 TCAATGGCTT CTGAGAAAAT GAAGATCCTG GAGCTTCCAT TTGCCAGTGG GACAATGAGC
 ATGTTGGTGC TGTTGCCTGA TGAAGTCTCA GGCCTTGAGC AGCTTGAGAG TATAATCAAC
 TTTGAAAAAC TGAATGAATG GACCAGTTCT AATGTTATGG AAGAGAGGAA GATCAAAGTG
 TACTTACCTC GCATGAAGAT GGAGGAAAAA TACAACCTCA CATCTGTCTT AATGGCTATG
 GGCATTACTG ACGTGTTTAG CTCTTCAGCC AATCTGTCTG GCATCTCCTC AGCAGAGAGC
 CTGAAGATAT CTCAGCTGT CCATGCAGCA CATGCAGAAA TCAATGAAGC AGGCAGAGAG
 GTGTTAGGCT CAGCAGAGGC TGGAGTGGAT GCTGCAAGCG TCTCTGAAGA ATTGAGGGCT
 GACCATCCAT TCCTCTTCTG TATCAAGCAC ATCGCAACCA ACGCCGTTCT CTTCTTTGGC
 AGATGTGTTT CCCCT

SEQ ID NO:20 (Chicken Ovalbumin Signal Sequence - shortened 50bp)
 ATG GGCTCCATCG GCGCAGCAAG CATGGAATTT TGTTTGATG TATTCAAGGA

SEQ ID NO:21 (Chicken Ovalbumin Signal Sequence - shortened 100bp)
 ATG GGCTCCATCG GCGCAGCAAG CATGGAATTT TGTTTGATG TATTCAAGGA GCTCAAAGTC
 CACCATGCCA ATGAGAACAT CTTCTACTGC CCCATTGCCA

SEQ ID NO:22 (vitellogenin targeting sequence).
 ATGAGGGGGATCATACTGGCATTAGTGCTCACCCCTGTAGGCAGCCAGAAGTTTGACATGGT

SEQ ID NO:23 (pro-insulin sequence)
 TTGTGAACCAACACCTGTGCGGCTCACACCTGGTGGAAAGCTCTCTACCTAGTGTGCGGGGAACGAGGC
 TTCTTCTACACACCAAGACCCGCGGAGGCAGAGGACCTGCAGGTGGGGCAGGTGGAGCTGGGCGGG
 GGCCCTGGTGCAGGCAGCCTGCAGCCCTTGGCCCTGGAGGGGTCCCTGCAGAAGCGTGGCATTGTGGAA
 CAATGCTGTACCAGCATCTGCTCCCTCTACCAGCTGGAGAACTCTGCAACTAG

SEQ ID NO:24 (p146 protein)
 KYKKALKKLAKLL

SEQ ID NO:25 (p146 coding sequence)
 AAATACAAAAAGCACTGAAAAAAGTGGCAAACTGCTG

SEQ ID NO:26 (spacer)
 (GPGG)_x

SEQ ID NO:27 (spacer)
 GPGGGPGGGPGG

SEQ ID NO:28 (spacer)
 GGGGSGGGSGGGG

SEQ ID NO:29 (spacer)

GGGGSGGGGSGGGGSGGGGS

SEQ ID NO:30 (repeat domain in TAG spacer sequence)
Pro Ala Asp Asp Ala

5

SEQ ID NO:31 (TAG spacer sequence)
Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp
Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp

10

SEQ ID NO:32 (gp41 epitope)
Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys Gly Trp Ile Gly Leu Leu

15

SEQ ID NO:33 (polynucleotide sequence encoding gp41 epitope)
Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr Cys Ile Leu Lys Gly
Ser Cys Gly Trp Ile Gly Leu Leu Asp Asp Asp Asp Lys

20

SEQ ID NO:34 (enterokinase cleavage site)
DDDDK

25

SEQ ID NO:35 (TAG sequence)
Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp
Ala Pro Ala Asp Asp Ala Pro Ala Asp Asp Ala Thr Thr Cys Ile Leu Lys Gly Ser Cys
Gly Trp Ile Gly Leu Leu Asp Asp Asp Asp Lys

30

SEQ ID NO:36 (altered transposase Hef forward primer)
ATCTCGAGACCATGTGTOAACTTGATATTTTACA3GATTCTCTTTACC

35

SEQ ID NO:37 (altered transposase Her reverse primer)
GATTGATCATTATCATAATTTCCCCAAAGCGTAACC

SEQ ID NO:38 (Xho I restriction site)
CTCGAG

40

SEQ ID NO:39 (Bcl I restriction site)
TGATCA

SEQ ID NO:40 (CMVf-NgoM IV primer)
TTGCCGGCATCAGATTGGCTAT

45

SEQ ID NO:41 (Syn-polyAr-BstE II primer)
AGAGGTCACCGGGTCAATTCTTCAGCACCTGGTA

50

55

SEQ ID NO:42 (pTnMod(Oval/ENT tag/Proins/PA)- Chicken)

5 CTGACGCGCC CTGTAGCGGC GCATTAAGCG CGGCGGGTGT GGTGGTTACG 50
 CGCAGCGTGA CCGCTACACT TGCCAGCGCC CTAGCGCCCG CTCCTTTCCG 100
 TTTCTTCCCT TCCTTTCTCG CCACGTTCCG CGGCATCAGA TTGGCTATTG 150
 GCCATTGCAT ACGTTGTATC CATATCATAA TATGTACATT TATATTGGCT 200
 CATGTCCAAC ATTACCGCCA TGTGACATT GATTATTGAC TAGTTATTAA 250
 TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA TGGAGTTCCG 300
 CGTTACATAA CTTACGGTAA ATGGCCCGCC TGGCTGACCG CCCAACGACC 350
 CCGGCCCATT GACGTCAATA ATGACGTATG TTCCCATAGT AACGCCAATA 400
 10 GGGACTTTCC ATTGACGTCA ATGGGTGGAG TATTTACGGT AAAGTGCCCA 450
 CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTACGCCC CCTATTGACG 500
 TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCCAGTA CATGACCTTA 550
 TGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA TCGCTATTAC 600
 CATGGTGTATG CGGTTTGGGC AGTACATCAA TGGGCGTGGA TAGCGGTTTG 650
 ACTCACGGGG ATTTCGAAGT CTCCACCCCA TTGACGTCAA TGGGAGTTTG 700
 15 TTTTGGCACC AAAATCAACG GGACTTTCCA AAATGTCGTA ACAACTCCGC 750
 CCCATTGACG CAAATGGGCG GTAGGCGTGT ACGGTGGGAG GTCTATATAA 800
 GCAGAGCTCG TTTAGTGAAC CGTCAGATCG CCTGGAGACG CCATCCACGC 850
 TGTTTTGACC TCCATAGAAG ACACCGGGAC CGATCCAGCC TCCGCGGCCG 900
 GGAACGGTGC ATTGGAACGC GGATTCCCGG TGCCAAGAGT GACGTAAGTA 950
 20 CCGCCTATAG ACTCTATAGG CACACCCCTT TGGCTCTTAT GCATGCTATA 1000
 CTGTTTTTGG CTTGGGGCCT ATACACCCCT GCTTCCTTAT GCTATAGGTG 1050
 ATGGTATAGC TTAGCCTATA GGTGTGGGTT ATTGACCATT ATTGACCACT 1100
 CCCCTATTGG TGACGATACT TTCCATTACT AATCCATAAC ATGGCTCTTT 1150
 GCCACAACCTA TCTCTATTGG CTATATGCCA ATACTCTGTC CTTGAGAGAC 1200
 TGACACGGAC TCTGTATTTT TACAGGATGG GGTCCCATTT ATTATTTACA 1250
 25 AATTACACATA TACAACAACG CCGTCCCCCG TGCCCCGAGT TTTTATTAAA 1300
 CATAGCGTGG GATCTCCACG CGAATCTCGG GTACGTGTTT CCGACATGGG 1350
 CTCTTCTCCG GTAGCGGCGG AGCTTCCACA TCCGAGCCCT GGTCCCATGC 1400
 CTCCAGCGGC TCATGGTCCG TCGGCAGCTC CTTGCTCCTA ACAGTGGAGG 1450
 CCAGACTTAG GCACAGCACA ATGCCACCA CCACGAGTGT GCCGCACAAG 1500
 30 GCGGTGGCGG TAGGGTATGT GTCTGAAAT GAGCGTGGAG ATTGGGCTCG 1550
 CACGGCTGAC GCAGATGGAA GACTTAAGGC AGCGGCAGAA GAAGATGCAG 1600
 GCAGCTGAGT TGTGTATTTC TGATAAGAGT CAGAGGTAAC TCCCGTTGCG 1650
 GTGCTGTAA CGGTGGAGGG CAGTGTAGTC TGAGCAGTAC TCGTTGCTGC 1700
 CGCGCGCGCC ACCAGACATA ATAGCTGACA GACTAACAGA CTGTTCTCTT 1750
 CCATGGGTCT TTTCTGCAGT CACCGTCGGA CCATGTGTGA ACTTGATATT 1800
 35 TTACATGATT CTCTTTACCA ATTCTGCCCC GAATTACACT TAAAACGACT 1850
 CAACAGCTTA ACGTTGGCTT GCCACGCATT ACTTGACTGT AAAACTCTCA 1900
 CTCTTACCGA ACTTGGCCGT AACCTGCCAA CCAAAGCGAG AACAAAACAT 1950
 AACATCAAAC GAATCGACCG ATTGTTAGGT AATCGTCACC TCCACAAAGA 2000
 GCGACTCGCT GTATACCGTT GGCATGCTAG CTTTATCTGT TCGGGAATAC 2050
 40 GATGCCCATT GTACTTGTG ACTGGTCTGA TATTCGTGAG CAAAAACGAC 2100
 TTATGGTATT GCGAGCTTCA GTCGCACTAC ACGGTCGTTT TGTACTCTT 2150
 TATGAGAAAG CGTTCCCGCT TTCAGAGCAA TGTTCAAAGA AAGCTCATGA 2200
 CCAATTCTTA GCGACCTTG CGAGCATTCT ACCGAGTAAC ACCACACCGC 2250
 TCATTGTGAG TGATGCTGGC TTAAAGTGC CATGGTATAA ATCCGTTGAG 2300
 AAGCTGGGTT GGTACTGGTT AAGTCGAGTA AGAGGAAAAG TACAATATGC 2350
 45 AGACCTAGGA GCGGAAACT GGAACCTAT CAGCAACTTA CATGATATGT 2400
 CATCTAGTCA CTCAAAGACT TTAGGCTATA AGAGGCTGAC TAAAAGCAAT 2450
 CCAATCTCAT GCCAAATTCT ATTGTATAAA TCTCGTCTA AAGGCCGAAA 2500
 AAATCAGCGC TCGACACGGA CTCATTGTCA CCACCCGTCA CCTAAAATCT 2550

5 ACTCAGCGTC GGCAAAGGAG CCATGGGTTT TAGCAACTAA CTTACCTGTT 2600
 GAAATTCGAA CACCCAAACA ACTTGTAAAT ATCTATTCGA AGCGAATGCA 2650
 GATTGAAGAA ACCCTCCGAG ACTTGAAAAG TCCTGCCTAC GGACTAGGCC 2700
 TACGCCATAG CCGAACGAGC AGCTCAGAGC GTTTTGATAT CATGCTGCTA 2750
 ATCGCCCTGA TGCTTCAACT AACATGTTGG CTTGCGGGCG TTCATGCTCA 2800
 GAAACAAGGT TGGGACAAGC ACTTCCAGGC TAACACAGTC AGAAATCGAA 2850
 ACGTACTCTC AACAGTTTCGC TTAGGCATGG AAGTTTTCGG GCATTCTGGC 2900
 TACACAATAA CAAGGGAAGA CTTACTCGTG GCTGCAACCC TACTAGCTCA 2950
 10 AAATTTATTC ACACATGGTT ACGCTTTGGG GAAATTATGA TAATGATCCA 3000
 GATCACTTCT GGCTAATAAA AGATCAGAGC TCTAGAGATC TGTGTGTTGG 3050
 TTTTTTGTGG ATCTGCTGTG CCTTCTAGTT GCCAGCCATC TGTGTGTTGC 3100
 CCCTCCCCCG TGCTTCTCTT GACCCTGGAA GGTGCCACTC CCACTGTCCT 3150
 TTCCTAATAA AATGAGGAAA TTGCATCGCA TTGTCTGAGT AGGTGTCATT 3200
 CTATTCTGGG GGGTGGGGTG GGGCAGCACA GCAAGGGGGA GGATTGGGAA 3250
 15 GACAATAGCA GGCATGCTGG GGATGCGGTG GGCTCTATGG GTACCTCTCT 3300
 CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CTCTCGGTAC CTCTCTCTCT 3350
 CTCTCTCTCT CTCTCTCTCT CTCTCTCTCT CCGTACCAGG TGCTGAAGAA 3400
 TTGACCCGGT GACCAAAGGT GCCTTTTATC ATCATTAA AAATAAAAAA 3450
 CAATTACTCA GTGCTGTGTA TAAGCAGCAA TTAATTATGA TTGATGCCTA 3500
 20 CATCACAACA AAACTGATT TAACAAATGG TTGCTCGCC TTAGAAAGTA 3550
 TATTGTGAACA TTATCTTGAT TATATTATTG ATAATAATAA AAACCTTATC 3600
 CCTATCCAAG AAGTGATGCC TATCATTGGT TGGAATGAAC TTGAAAAAAA 3650
 TTAGCCTTGA ATACATTACT GGTAAGGTAA ACGCCATTGT CAGCAAATG 3700
 ATCCAAGAGA ACCAACTTAA AGCTTTCCTG ACGGAATGTT AATTCTCGTT 3750
 GACCTTGAGC ACTGATGAAT CCCCTAATGA TTTTGGTAAA AATCATTAA 3800
 25 TTAAGGTGGA TACACATCTT GTCATATGAT CCCGGTAATG TGAGTTAGCT 3850
 CACTCATTAG GCACCCAGG CTTTACACTT TATGCTTCCG GCTCGTATGT 3900
 TGTGTGGAAT TGTGAGCGGA TAACAATTTT ACACAGGAAA CAGCTATGAC 3950
 CATGATTACG CCAAGCGCGC AATTAACCTT CACTAAAGGG AACAAAAGCT 4000
 GGAGCTCCAC CGCGGTGGCG GCCGCTCTAG AACTAGTGA TCCCCGGGG 4050
 30 AGGTCAAGAT GGTTCCTTTA CTGTTTGTCA ATTCTATTAT TTCAATACAG 4100
 AACAAAGCT TCTATAACTG AAATATATTT GCTATTGTAT ATTATGATTG 4150
 TCCCTCGAAC CATGAACACT CCTCCAGCTG AATTTACAA TTCTCTGTC 4200
 ATCTGCCAGG CCATTAAGTT ATTCATGGAA GATCTTTGAG GAACACTGCA 4250
 AGTTCATATC ATAAACACAT TTGAAATGA GTATTGTTT GCATTGTATG 4300
 GAGCTATGTT TTGCTGTATC CTCAGAAAAA AAGTTTGTGA TAAAGCAATC 4350
 35 ACACCCATAA AAAGATAGAT TTAAATATTC CAGCTATAGG AAAGAAAGTG 4400
 CGTCTGCTCT TCACTCTAGT CTCAGTTGGC TCCTTCACAT GCATGCTTCT 4450
 TTATTTCTCC TATTTTGTCA AGAAAAATAT AGGTACAGTC TTGTCTCAC 4500
 TTATGTCTG CCTAGCATGG CTCAGATGCA CGTTGTAGAT ACAAGAAGGA 4550
 TCAAAATGAAA CAGACTTCTG GTCTGTTACT ACAACCATAG TAATAAGCAC 4600
 ACTAACTAAT AATTGCTAAT TATGTTTTC ATCTCTAAGG TTCCCACTT 4650
 40 TTTCTGTTTT CTTAAAGATC CCATTATCTG GTTGTAATG AAGCTCAATG 4700
 GAACATGAGC AATATTTCCT AGTCTTCTCT CCCATCCAAC AGTCCTGATG 4750
 GATTAGCAGA ACAGGCAGAA AACACATTGT TACCCAGAAT TAAAACTAA 4800
 TATTTGCTCT CCATTCAATC CAAAATGGAC CTATTGAAAC TAAAATCTAA 4850
 CCCAATCCCA TTAAATGATT TCTATGGCGT CAAAGGTCAA ACTTCTGAAG 4900
 45 GGAACCTGTG GGTGGGTCAC AATTGAGGCT ATATATTTCC CAGGGCTCAG 4950
 CGGATCCATG GGCTCCATCG GCGCAGCAAG CATGGAATTT TGTITTGATG 5000
 TATTCAAGGA GCTCAAAGTC CACCATGCCA ATGAGAACAT CTCTACTGTC 5050
 CCCATTGCCA TCATGTCAGC TCTAGCCATG GTATACCTGG GTGCAAAAGA 5100
 CAGCACCAGG ACACAGATAA ATAAGGTTGT TCGCTTTGAT AAACCTCCAG 5150
 GATTGCGAGA CAGTATTGAA GCTCAGTGTG GCACATCTGT AAACGTTTAC 5200
 50 TCTTCACTTA GAGACATCCT CAACCAATC ACCAAACCA ATGATGTTTA 5250
 TTGTTTCAGC CTTGCCAGTA GACTTTATGC TGAAGAGAGA TACCAATCC 5300
 TGCCAGAATA CTTGCAAGTGT GTGAAGGAAC TGTATAGAGG AGGCTTGGAA 5350
 CCTATCAACT TTCAAACAGC TGCAGATCAA GCCAGAGAGC TCATCAATTC 5400
 CTGGGTAGAA AGTCAGACAA ATGGAATTAT CAGAAATGTC CTTAGCCAA 5450
 GCTCCGTGGA TTCTCAAAC GCAATGGTTC TGGTTAATGC CATTGTCTTC 5500
 55 AAAGGACTGT GGGAGAAAAC ATTTAAGGAT GAAGACACAC AAGCAATGCC 5550
 TTTCAGAGTG ACTGAGCAAG AAAGCAAACC TGTGCAGATG ATGTACCAGA 5600

	TTGGTTTATT	TAGAGTGGCA	TCAATGGCTT	CTGAGAAAAT	GAAGATCCTG	5650
	GAGCTTCCAT	TTGCCAGTGG	GACAATGAGC	ATGTTGGTGC	TGTTGCCTGA	5700
5	TGAAGTCTCA	GGCCTTGAGC	AGCTTGAGAG	TATAATCAAC	TTTGAAAAAC	5750
	TGACTGAATG	GACCAGTTCT	AATGTTATGG	AAGAGAGGAA	GATCAAAGTG	5800
	TACTTACCTC	GCATGAAGAT	GGAGGAAAAA	TACAACTCA	CATCTGTCTT	5850
	AATGGCTATG	GGCATTACTG	ACGTGTTTAG	CTCTTCAGCC	AATCTGTCTG	5900
	GCATCTCCTC	AGCAGAGAGC	CTGAAGATAT	CTCAAGCTGT	CCATGCAGCA	5950
	CATGCAGAAA	TCAATGAAGC	AGGCAGAGAG	GTGGTAGGGT	CAGCAGAGGC	6000
10	TGGAGTGGAT	GCTGCAAGCG	TCTCTGAAGA	ATTTAGGGCT	GACCATCCAT	6050
	TCCTCTTCTG	TATCAAGCAC	ATCGCAACCA	ACGCCGTTCT	CTTCTTTGGC	6100
	AGATGTGTTT	CCCTCCGCG	GCCAGCAGAT	GACGCACCAG	CAGATGACGC	6150
	ACCAGCAGAT	GACGCACCAG	CAGATGACGC	ACCAGCAGAT	GACGCACCAG	6200
	CAGATGACGC	AACAACATGT	ATCCTGAAAG	GCTCTGTGG	CTGGATCGGC	6250
	CTGCTGGATG	ACCATGACAA	ATTGTGTAAC	CAACACCTGT	GCGGCTCACA	6300
15	CCTGGTGGAA	GCTCTCTACC	TAGTGTGCGG	GGAACGAGGC	TTCTTCTACA	6350
	CACCCAAGAC	CCGCCGGGAG	GCAGAGGACC	TGCAGGTGGG	GCAGGTGGAG	6400
	CTGGGCGGGG	GCCCTGGTGC	AGGCAGCCTG	CAGCCCTTGG	CCCTGGAGGG	6450
	GTCCCTGCAG	AAGCGTGGCA	TTGTGGAACA	ATGCTGTACC	AGCATCTGCT	6500
	CCCTCTACCA	GCTGGAGAAC	TACTGCAACT	AGGGCGCCTG	GATCCAGATC	6550
20	ACTTCTGGCT	AATAAAAGAT	CAGAGCTCTA	GAGATCTGTG	TGTTGGTITT	6600
	TTGTGGATCT	GCTGTGCCCT	CTAGTTGCCA	GCCATCTGTT	GTTTGCCCTT	6650
	CCCCCGTGCC	TTCTTTGACC	CTGGAAGGTG	CCACTCCAC	TGTCCTTTCC	6700
	TAATAAAATG	AGGAAATTGC	ATCGCATTGT	CTGAGTAGGT	GTCATTCTAT	6750
	TCTGGGGGGT	GGGGTGGGGC	AGCACAGCAA	GGGGGAGGAT	TGGGAAGACA	6800
	ATAGCAGGCA	TGCTGGGGAT	GCGGTGGGCT	CTATGGGTAC	CTCTCTCTCT	6850
25	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CGGTACCTCT	CTCGAGGGGG	6900
	GGCCCCGTAC	CCAATTCCGC	CTATAGTGAG	TGCTATTACG	GCGCTCACT	6950
	GGCCGTCGTT	TTACAACGTC	GTGACTGGGA	AAACCCTGGC	GTTACCCAAC	7000
	TTAATCGCCT	TGCAGCACAT	CCCCCTTTCG	CCAGCTGGCG	TAATAGCGAA	7050
	GAGGCCCGCA	CCGATCGCCC	TTCCCAACAG	TTGCCGAGCC	TGAATGGCGA	7100
	ATGGAAATTG	TAAGCGTTAA	TATTTTGTTA	AAATTCGCGT	TAAATTTTGT	7150
30	TTAAATCAGC	TCATTTTTTA	ACCAATAGGC	CGAAATCGGC	AAAATCCCTT	7200
	ATAAATCAAA	AGAATAGACC	GAGATAGGGT	TGAGTGTGT	TCCAGTTTGG	7250
	AACAAGAGTC	CACTATTAAA	GAACGTGGAC	TCCAACGTCA	AAGGGCGAAA	7300
	AACCGTCTAT	CAGGGCGATG	GCCCCACTACT	CCGGGATCAT	ATGACAGAT	7350
	GTGTATCCAC	CITAACTTAA	TGATTTTAC	CAAAATCATT	AGGGGATTCA	7400
35	TCAGTGTCTA	GGGTCAACGA	GAATTAACAT	TCCGTACAGG	AAGCTTATGA	7450
	TGATGATGTG	CTTAAAACT	TACTCAATGG	CTGGTTATGC	ATATCGCAAT	7500
	ACATGCGAAA	AACCTAAAAG	AGCTTGCCGA	TAAAAAAGGC	CAATTTATTG	7550
	CTATTTACCG	CGGCTTTTAA	TTGAGCTTGA	AAGATAAATA	AAATAGATAG	7600
	GTTTTATTGT	AAGCTAAATC	TTCTTTATCG	TAAAAAATGC	CCTCTTGGGT	7650
	TATCAAGAGG	GTCAATATAT	TTGCGGGAAT	AACATCATTT	GGTGACGAAA	7700
40	TAACTAAGCA	CTTGTCTCCT	GTTTACTCCC	CTGAGCTTGA	GGGGTTAACA	7750
	TGAAGGTGAT	CGATAGCAGG	ATAATAATAC	AGTAAAACGC	TAAACCAATA	7800
	ATCCAAATCC	AGCCATCCCA	AATTGGTAGT	GAATGATTAT	AAATAACAGC	7850
	AAACAGTAAT	GGGCCAATAA	CACCGGTTGC	ATTGGTAAGG	CTCACCATA	7900
	ATCCCTGTAA	AGCACCTTGC	TGATGACTCT	TGTTTGGAT	AGACATCACT	7950
45	CCCTGTAAATG	CAGGTAAAGC	GATCCCAACA	CCAGCCAATA	AAATTAAAAC	8000
	AGGGAAAACT	AACCAACCTT	CAGATATAAA	CGCTAAAAAG	GCAATGCAC	8050
	TACTATCTGC	AATAAATCCG	AGCAGTACTG	CCGTTTTTTC	GCCCCATTTA	8100
	GTGGCTATTC	TTCTTGCCAC	AAAGGCTTGG	AATACTGAGT	GTAAAAGACC	8150
	AAGACCCGCT	AATGAAAAGC	CAACCATCAT	GCTATTCCAT	CCAAAACGAT	8200
	TTTCGGTAAA	TAGCACCCAC	ACCGTTGCGG	GAATTTGGCC	TATCAATTGC	8250
50	GCTGAAAAAT	AAATAATCAA	CAAAATGGCA	TCGTTTTTAA	TAAAGTGATG	8300
	TATACCGAAT	TCAGCTTTTG	TTCCCTTTAG	TGAGGGTTAA	TTGCGCGCTT	8350
	GGCGTAAATCA	TGGTCATAGC	TGTTTCCTGT	GTGAAATGT	TATCCGCTCA	8400
	CAATTCCACA	CAACATACGA	GCCGGAAGCA	TAAAGTGTA	AGCCTGGGGT	8450
	GCCTAATGAG	TGAGCTAACT	CACATTAATT	GCGTTGCGCT	CACTGCCCGC	8500
	TTTCCAGTCG	GGAAACCTGT	CGTGCCAGCT	GCATTAAATGA	ATCGGCCAAC	8550
55	GCGCGGGGAG	AGGCGGTTTG	CGTATTGGGC	GCTCTTCGCG	TTCTCGCTC	8600
	ACTGACTCGC	TGCGCTCGGT	CGTTCGGCTG	CGGCGAGCGG	TATCAGCTCA	8650

CTCAAAGGCG GTAATACGGT TATCCACAGA ATCAGGGGAT AACGCAGGAA 8700
 AGAACATGTG AGCAAAAGGC CAGCAAAAGG CCAGGAACCG TAAAAAGGCC 8750
 GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC CCCCTGACG AGCATCACAA 8800
 5 AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCCGACAGGA CTATAAAGAT 8850
 ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG TCGCTCTCC TGTTCGACC 8900
 CTGCCGCTTA CCGGATACCT GTCCGCCTTT CTCCCTTCGG GAAGCGTGGC 8950
 GCTTTCTCAT AGCTCACGCT GTAGGTATCT CAGTTCGGTG TAGGTCGTT 9000
 GCTCCAAGCT GGGCTGTGTG CACGAACCCC CCGTTCAGCC CGACCGCTGC 9050
 10 GCCTTATCCG GTAACATATG TCTTGAGTCC AACCCGGTAA GACACGACTT 9100
 ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA GCGAGGTATG 9150
 TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA CGGCTACACT 9200
 AGAAGGACAG TATTGTGAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG 9250
 AAAAAGAGTT GGTAGCTCTT GATCCGGCAA ACAAAACCACC GCTGGTAGCG 9300
 GTGGTTTTTT TGTGTGCAAG CAGCAGATTA CGCGCAGAAA AAAAGGATCT 9350
 15 CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA 9400
 AAACCTCAGT TAAGGGATT TGGTCATGAG ATTATCAAAA AGGATCTTCA 9450
 CCTAGATCCT TTAAATTA AAATGAAGTT TTAATCAAT CTAAAGTATA 9500
 TATGAGTAAA CTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC 9550
 TATCTCAGCG ATCTGTCTAT TTCGTTTATC CATAGTTGCC TGACTCCCCG 9600
 TCGTGTAGAT AACTACGATA CGGGAGGGCT TACCATCTGG CCCAGTGCCT 9650
 20 GCAATGATAC CGCGAGACCC ACGCTCACCG GCTCCAGATT TATCAGCAAT 9700
 AAACCAGCCA GCGGAAGGG CCGAGCGCAG AAGTGGTCTT GCAACTTTAT 9750
 CCGCTCCAT CCAGTCTATT AATTGTTGCC GGGAACTAG AGTAAGTAGT 9800
 TCGCCAGTTA ATAGTTTGG CAACGTTGTT GCCATTGCTA CAGGCATCGT 9850
 GGTGTACGCG TCGTCTTGT GTATGGCTTC ATTACGCTCC GGTTCCTAAC 9900
 25 GATCAAGGCG AGTTACATGA TCCCCATGT TGTGCAAAAA AGCGGTTAGC 9950
 TCCTTCGGTC CTCCGATCGT TGTGAGAAGT AAGTTGGCCG CAGTGTATC 10000
 ACTCATGGTT ATGGCAGCAC TGCATAATC TCTTACTGTC ATGCCATCCG 10050
 TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA 10100
 TAGTGTATGC GCGCAGCGAG TTGCTCTTGC CCGGCGTCAA TACGGGATAA 10150
 TACCGCGCCA CATAGCAGAA CTTTAAAGT GCTCATCATT GGAAAACGTT 10200
 30 CTTCGGGGCG AAAACTCTCA AGGATCTTAC CGCTGTTGAG ATCCAGTTCG 10250
 ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTAATTTTAC 10300
 CAGCGTTTCT GGGTGAGCAA AAACAGGAAG GCAAAATGCC GCAAAAAGG 10350
 GAATAAGGGC GACACGAAA TGTGAATAC TCATACTCTT CCTTTTCAA 10400
 TATTATTGAA GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT 10450
 35 TGAATGTATT TAGAAAAATA AACAAATAGG GGTCCGCGC ACATTTCCCC 10500
 GAAAAGTGCC AC 10512

40 SEQ ID NO:43 (pTnMOD (CMV-CHOVg-ent-ProInsulin-synPA))

1 ctgaagcgcc ctgtagcgcc gcattaagcg cggcggtgt ggtggttac cgcagcgtga
 61 ccgctacact tgccagcgcc ctacgcccgc ctcctttcgc tttcttccct tctttctcg
 121 ccacgttcgc cggcatcaga ttggtattg gccattgcat acgttgtatc catatcataa
 181 tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac
 241 tagttattaa tagtaataa ttacgggggc attagtcat agcccatata tggagttccg
 301 cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt
 361 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca
 421 atgggtggag tatttacggt aaactgcccc cttggcagta catcaagtgt atcatatgcc
 481 aagtaagccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta
 541 catgacctta tgggactttc ctacttggca gtacatctac gtatttagtca tgcctattac
 601 catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcaagggg
 661 atttccaagt ctcaccccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacy
 721 ggactttcca aaatgtcgta acaactccgc cccattgacg caaatggggc gtaggcgtgt
 781 acggtgggag gtctatatga gcagagctcg tttagtgaac cgtcagatcg cctggagacg
 841 ccattccacg tgttttgacc tccatagaag acaccgggac cgtaccagcc tcccgggcgg
 901 ggaacggtgc attggaacgc ggattccccg tgccaagagt gacgtaagta ccgcctatag
 961 actctatagg cacacccctt tggctcttat gcatgctata ctgtttttgg cttggggcct
 1021 atacaccccc gtttctctat gctatagggt atggtatagc ttagecctata ggtgtggggt
 1081 attgaccatt attgaccact cccctattgg tgacgatact ttccattact aatccataac
 1141 atggctcttt gccacaacta tctctattgg ctatatgcca atactctgtc cttcagagac
 1201 tgacacggac tctgtatttt tacaggatgg ggtcccatth attatttaca aattcacata

5

10

15

20

25

30

35

40

45

50

55

1261 tacaacaacg ccgtcccccg tgcgcgcagt ttttattaaa catagcgtgg gatctccacg
 1321 cgaatctcgg gtacgtgttc cggacatggg ctcttctcgg gtacgaggcg agcttccaca
 1381 tccgagccct ggtcccatgc ctccagcggc tcatggtcgc tcggcagctc ctgtctccta
 1441 acagtgaggg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag
 1501 gccgtggcgg tagggtagt gtctgaaaat gagcgtggag attgggctcg cacggctgac
 1561 gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgatttc
 1621 tgataagagt cagaggtaac tcccgttgcg gtgctgttaa cgggtggaggg cagtgtagtc
 1681 tgagcagtac tcgttgctgc cgcgcgcgcc accagacata atagctgaca gactaacaga
 1741 ctgttccttt ccatgggtct tttctgcagt caccgtcgga ccatgtgtga acttgatatt
 1801 ttacatgatt ctctttacca attctgcccc gaattacact taaaacgact caacagctta
 1861 acgttggcct gccacgcatt acttgactgt aaaactctca ctcttaccga acttggcogt
 1921 aacctgcca ccaagcgag aacaaaacat aacatcaaac gaatcgaccg attggttaggt
 1981 aatggtcacc tccacaaaga gogactcgct gtatacogtt ggcattgctag ctttatctgt
 2041 tcgggcaata cgaagcccat tgtactgttt gactggctgt atattcgtga gcaaaaaaga
 2101 cttatggtat tgcgagcttc agtcgcacta cacggctcgt ctgttactct ttatgagaaa
 2161 gogttcccg cttcagagca atgttcaaa aaagctcag accaatttct agcgcagctt
 2221 gcgagcattc taccgagtaa caccacaccg ctcatgtca gtgatgctgg ctttaaagt
 2281 ccatgggtata aatccgttga gaagctgggt tggtagctgt taagtcgagt aagaggaaaa
 2341 gtacaatatg cagacctagg agcggaaaac tggaaaccta tcagcaactt acatgatatg
 2401 tcatctagtc actcaaagac ttttaggtat aagaggctga ctaaaagcaa tccaatctca
 2461 tgccaaattc tattgtataa atctcgctct aaaggccgaa aaaatcagcg ctcgacacgg
 2521 actcattgtc accaccgcgc acctaaaac tactcagcgt cggcaaaagg gccatgggtt
 2581 ctagcaacta acttacctgt tgaatttcga acacccaaac aacttgtaa tatctattcg
 2641 aagcgaatgc agattgaaga aaccttcga gacttgaata gtccctgccta cggactaggc
 2701 ctacgcata gccgaacgag cagctcagag cgttttgata tcatgctgct aatcgccctg
 2761 atgcttcaac taacatgttg gcttgcgggc gttcatgtc agaaacaagg ttgggacaag
 2821 cacttccagg ctaacacagt cagaaatcga aacgtactct caacagttcg cttaggcatg
 2881 gaagttttgc ggcattctgg ctacacaata acaagggaag acttactcgt ggtcgcaacc
 2941 ctactagctc aaaattctatt cacacatggt tacgctttgg ggaatttatg ataagtatcc
 3001 agatcacttc tggctaataa aagatcagag ctctagagat ctgtgtgttg gttttttgtg
 3061 gatctgtgt gccttctagt tgcagcccat ctgttgtttg cccctcccc gtgccttctc
 3121 tgacctggga aggtgccact cccactgtcc tttcctaata aaatgaggaa attgcacgc
 3181 attgtctgag taggtgtcat tctattctgg ggggtggggg ggggcagcac agcaaggggg
 3241 aggattggga agacaatagc aggcattgct gggatgcggg gggctctatg ggtacctctc
 3301 tctctctctc tctctctctc tctctctctc tctctcggtg cctctctctc tctctctctc
 3361 tctctctctc tctctctctc tcggtaaccg gtgctgaaga attgaccgg tgaccaaaagg
 3421 tgccttttat catcacttta aaaataaaaa acaattactc agtgccctgtt ataagcagca
 3481 attaatatg attgatgcct acatcacaac aaaaactgat ttaacaaatg gttggctctgc
 3541 cttagaaagt atatttgaac attatcttga ttatatatt gataataata aaaccttat
 3601 cctatccaa gaagtgtatg ctatcattgg ttggaatgaa cttgaaaaaa attgaccttg
 3661 aatacattac tggtaaggta aacgccattg tcagcaaat gatccaagag aaccaactta
 3721 aagcttctct gacggaatgt taattctcgt tgacctgag cactgatgaa tccccaatg
 3781 attttggtaa aaatcattaa gctaagggtg atacacatct tgtcatatga tcccggtaat
 3841 gtgagttagc tcaactcatta ggcaccccg gctttacact ttatgcttcc ggtcgtatg
 3901 ttgtgtggaa ttgtgagcgg ataacaattt cacacaggaa acagctatga ccatgattac
 3961 gccaaagcgg caattaaccc tcaataaagg gaacaaaagc tggagctcca ccgcggtggc
 4021 ggcgcgtcta gaactagtgg atccccggg catcagattg gctattggcc attgcatacg
 4081 ttgtatccat atcataatat gtacatttat attggctcat gtccaacatt accgccatgt
 4141 tgacattgat tattgactag ttattaatag taatcaatta cggggtcatt agttcatagc
 4201 ccatatatgg agttccgcgt tacataaact acggtaaatg gccgcctcg ctgaccgccc
 4261 aacgaccccc gccattgac gtcaataatg acgtatgttc ccatagtaac gccaataggg
 4321 actttccatt gacgtcaatg ggtggagtat ttacggtaaa ctgccactt ggcagtagat
 4381 caagtgtatc atatgccaa gacgccccct attgacgtca atgacggtaa atggcccgcc
 4441 tggcattatg cccagtacat gaccttatgg gactttccta cttggcagta catctacgta
 4501 ttagtcatcg ctattacat ggtgatgcgg ttttggcagt acatcaatgg gcgtggatag
 4561 cggtttgact caggggatt tccaagtctc caccctattg acgtcaatgg gagtttgtt
 4621 tggcaccaaa atcaacggga ctttccaaa tgtcgtaaca actccgcccc attgacgcaa
 4681 atgggcggta ggcgtgtacg gtgggaggtc tatataagca gagctcgtt agtgaccgt
 4741 cagatcgccct ggagacgcca tccacgtgt tttgacctcc atagaagaca cggggacgga
 4801 tccagcctcc ggcgcggga acggtgcatt ggaacgcgga ttccccgtgc caagagtac
 4861 gtaagtaccg cctatagact ctataggcac acccctttgg ctcttatgca tgctatactg
 4921 tttttgctt ggggcctata cccccgct tccctatgct atagggtatg gtagactta
 4981 gccataggt gtgggttatt gacctattt gacctctcc ctattgggtg cgtactttc
 5041 cattactaat ccataacatg gctctttgcc acaactatct ctattggcta tatgccaata
 5101 ctctgtcctt cagagactga cagcgactct gtatttttac aggatggggg cccatttatt
 5161 atttacaat tcacatatc aacaacgcgg tccccgtgc ccgagtttt tatataacat
 5221 agcgtgggat ctccacgcga atctcgggta cgtgttccgg acatgggctc ttctcggta
 5281 ggcgcggagc ttccacatcc gagccctggt ccatgcctc cagcggtca tggctcgtcg

5

10

15

20

25

30

35

40

45

50

55

5341 gcagctcctt gctcctaaca gtggaggcca gacttaggca cagcacaatg cccaccacca
 5401 ccagtgtgcc gcacaaggcc gtggcggtag ggtatgtgtc tgaaaatgag cgtggagatt
 5461 gggctcgac ggctgacgca gatggaagac ttaaggcagc ggcagaagaa gatgcaggca
 5521 gctgagttgt tttattctga taagagtcag aggttaactcc cgttgcggtg ctgttaacgg
 5581 tggagggcag tttagtctga gcagtaactc ttgctgcccgc gcgcgcacc agacataata
 5641 gctgacagac taacagactg ttctcttcca tgggtctttt ctgcagtcac cgtcggatca
 5701 atggggtcca tgggtgcagc aagcatggaa tttgtttttg atgtattcaa ggagctcaaa
 5761 gtccaccatg ccaatgagaa catcttctac tgccccattg ccatcatgtc agctctagcc
 5821 atgggtatacc tgggtgcaaa agacagcacc aggcacacaa taaataagggt tgttcgcttt
 5881 gataaaacttc caggattcgg agacagtatt gaagctcagt gtggcacatc tgtaaaagtt
 5941 cactcttcac tttagagacat cctcaacca atcaccaaac caaatgatgt ttattcgttc
 6001 agccttgcca gttagacttta tgctgaagag agatacccaa tcttgccaga atacttgca
 6061 tgtgtgaagg aactgtatag agggagcttg gaacctatca actttcaaac agctgcagat
 6121 caagccagag agctcatcaa ttcttggtga gaaagtcaga caaatggaat tatcagaat
 6181 gtccttcagc caagctcgtt ggattctcaa actgcaatgg ttctggttaa tgccattgtc
 6241 ttcaaggac tggggagaa agcatttaag gatgaagaca cacaagcaat gctttcaga
 6301 gtgactgagc aagaaagcaa acctgtgcag atgatgtacc agattgggtt atttagagt
 6361 gcatcaatgg cttctgagaa aatgaagatc ctggagcttc catttgccag tgggacaatg
 6421 agcatgttgg tgggttgcc tgatgaagtc tcaggccttg agcagcttga gagtataatc
 6481 aactttgaaa aactgactga atggaccagt tctaattgta tgggaagagag aagatcaaa
 6541 tgtacttacc tgcgatgaag atggaggaaa aatacaacct cacatctgtc ttaattggtc
 6601 tgggcattac tgacgtgttt agctcttcag ccaatctgtc tggcatctcc tcagcagaga
 6661 gcttgagat atctcaagct gtccatgcag cacatgcaga aatcaatgaa gcaggcagag
 6721 aggtgttagg gtcagcagag gctggagtg atgtgtcaag cgtctctgaa gaatttaggg
 6781 ctgaccatcc attctcttc tgtatcaagc acatgcgaac caagcagctt ctctctctt
 6841 ggcagatgtg tttcccgagg ccagcagatg acgcaccagc agatgacgca ccagcagatg
 6901 acgcaccagc agatgacgca ccagcagatg acgcacaac atgtatcctg aaaggctctt
 6961 gtggctggat cggcctgctg gatgacgatg acaaatgtt gaaccaacac ctgtgaggct
 7021 cacactgggt ggaagctctc tacctagtgt gcggggaacg aggtctcttc tacacacca
 7081 agaccgcgcg ggaggcagag gacctgcagg tggggcagggt ggagctgggc gggggccctg
 7141 gtgcaggcag cctgcagccc ttggcccttg aggggtccct gcagaagcgt ggcattgtgg
 7201 aacaatgctg taccagcacc tgcctccctc accagctgga gaactactgc aactaggggc
 7261 cctaaagggc gaattatcgc ggccgctcta gaccaggcgc ctggatccag atcactctg
 7321 gctaataaaa gatcagagct cttagagatc gtgtgttgggt ttttgtgga tctgctgtg
 7381 cttctagtgt ccagccatct gttgtttgcc cctccccgt gccttccttg accctggaag
 7441 gtgccactcc cactgtcctt tccataataa atgaggaaat tgcacgcat tgcctgagta
 7501 ggtgtcattc tttctggggg ggtgggggtg ggcagcacag caagggggag gatgggaag
 7561 acaatagcag gcatgctggg gatgcgggtg gctctatggg tacctctctc tctctctctc
 7621 tctctctcac tctctctctc tctcggtacc tctctctgag gggggggccc gtaccacaa
 7681 cgcctatag ttagtctgtat tacgcgcgct cactggcgtt cgttttaca cgtcgtgact
 7741 gggaaaacc tggcgttacc caacttaac gccttgagc acatccccct ttgcgcagct
 7801 ggcgttaatg cgaagaggcc cgcaccgatc gcccttccca acagtgtgag agcctgaatg
 7861 gcgaatggaa attgtaagcg ttaatatatt gttaaaatt gcgttaaat tttgttaaat
 7921 cagctcattt tttaaccaat agggcgaat cggcaaaat cctataaat caaaagata
 7981 gaccgagata ggggtgagtg ttgttccagt ttggaacaag agtcactat taaagaacgt
 8041 ggaactcaac gtcaaaaggc gaaaaaccgt ctatcagggc gatggccac tactccggga
 8101 tcatatgaca agatgtgtat ccaccttaac ttaatgattt ttacaaaaat cattagggga
 8161 ttcacagtg ctcagggtca acgagaatta acattccgtc aggaagagct atgatgatga
 8221 tgtgcttaaa aacttactca atggctggtt atgcataatg caatacatgc gaaaaacct
 8281 aaagagcttg ccgataaaaa agggcaattt attgtattt accgaggctt tttattgagc
 8341 ttgaaagata aataaaatag ataggtttta ttgaaagcta aatcttctt atcgtaaaaa
 8401 atgcccctct ggggtatcaa gagggtcatt atatttcgag gaataacatc atttgggtgac
 8461 gaaataacta agcacttgtc tctgttttac tccccgagc ttgaggggtt aacatgaagg
 8521 tcatcgatag caggataata atacagtaaa acgctaaacc aataatcaa atccagccat
 8581 cccaaattgg tagtgaatga ttataaataa cagcaaacag taatggggca ataacacgg
 8641 ttgcatgtgt aaggctcacc aataatccct gtaaagcacc ttgctgatga ctctttgtt
 8701 ggaatagacat cactccctgt aatgcaggt aagcgatccc accaccagcc aataaaatta
 8761 aaacagggaa aactaaccac ccttcagata taaacgctaa aaaggcaaat gcactactat
 8821 ctgcaataaa tccgagcagt actgcccgtt ttgcgcccac ttagtgggta tctctctgc
 8881 cacaaggct tggaaactcg agtgtaaaag accaagaccc gtaatgaaaa gccaacatc
 8941 atgctattca tcatcacgat ttctgttaata gcaccacac gtgctggatt ggtatcaat
 9001 gcgctgaat aataatcaac aaatggcatc gttaaataag tgatgtatc cgtacgctt
 9061 tgttccctt tagtgaggtt taattgcgcy ctggcgtaa tcatgggtcat agctgttctc
 9121 tgtgtgaaat tgttatccgc tcacaattcc acacaacata cgagccggaa gcataaagt
 9181 taaagccttg ggtgcctaat gaggtagcta actcacatta attgcgttgc gctcactgac
 9241 cgctttccag tgggaaacc tgcgtgcca gctgcattaa tgaatcgcc aacgcgggg
 9301 gagaggcgtt ttgcgtattg ggcgctcttc cgttctctc ctcactgact cgtgctgctc
 9361 ggtcgttcgg ctgcggcgag cggatcagc tcaactcaag gcggttaaac ggttatccac

5
10
15
20
25

9421 agaatcaggg gataacgcag gaaagaacat gtgagcaaaa ggccagcaaa aggccaggaa
 9481 cegtaaaaaag gccgcgttgc tggcggtttt ccataggctc cggccccctg acgagcatca
 9541 caaaaaatcga cgctcaagtc agaggtggcg aaaccgcaca ggactataaa gataccaggc
 9601 gtttccccctt ggaagctccc tctgctgctc tctgttccg accctgccgc ttaccggata
 9661 cctgtccgcc tttctccctt cgggaagcgt ggcgctttct catagctcac gctgtaggta
 9721 tctcagttcg gtgtaggtcg ttcgtccaa gctgggctgt gtgcacgaac ccccggttca
 9781 gcccgaccgc tgcgccttat ccggttaacta tctgtcttag tccaaccgg taagacacga
 9841 cttatcgcca ctggcagcag cactgtgtaa caggattagc agagcgaggt atgtaggcgg
 9901 tgctacagag ttcttgaagt ggtggcctaa ctacggctac actagaagga cagtatttgg
 9961 tatctgcgct ctgctgaagc cagtacctt cggaaaaaga gttggtagct cttgatccgg
 10021 caaacaacc accgctggta gcggtggtt tttgtttgc aagcagcaga ttacgcgcag
 10081 aaaaaaaggga tctcaagaag atcctttgat ctttctacg ggtctgcag ctcagtggaa
 10141 cgaaaactca cgttaaggga ttttggtcat gagattatca aaaaggatct tcacctagat
 10201 ccttttaaat taataatgaa gttttaaatc aatctaaagt atatatgagt aaacttggtc
 10261 tgacagttac caatgcttaa tcastgagge acctatctca gcgatctgtc tatttcgttc
 10321 atccatagtt gcctgactcc cgtcgtgta gataactacg atacgggagg gcttaccatc
 10381 tggccccagt gctgcaatga taccgcgaga cccacgctca ccggtccag atttatcagc
 10441 aataaaccag ccagccggaa gggccgagcg cagaagtgtt cctgcaactt tatccgctc
 10501 catccagtct attaattggt gccgggaagc tagagtaagt agttcgccag ttaatagttt
 10561 gcgcaacgtt gttgccattg ctacaggcat cgtggtgtca cgtcgtcgtc ttggtatggc
 10621 ttcattcagc tcoggttccc aacgatcaag gcgagttaca tgatccccc tgttgtgcaa
 10681 aaaagcgggt agctccttcg gtctcctgat cgttgcaga agtaagtgg ccgagtggt
 10741 atcactcatg gttatggcag cactgcataa ttctcttact gtcatgccat ccgtaagatg
 10801 cttttctgtg actggtgagt actcaaccaa gtcattctga gaatagtga tgcggcgacc
 10861 gagtgtctct tgcccggcgt caatacggga taataccgag ccacatagca gaactttaa
 10921 agtgcctatc attggaaaac gttcttcggg gcgaaaactc tcaaggatct taccgctgtt
 10981 gagatccagt ccatgttaac ccactcgtgc acccaactga tcttcagcat cttttacttt
 11041 caccagcgtt tctgggtgag caaaaacagg aaggcaaat gcgcgaaaa aggggaataag
 11101 ggcgacacgg aaatgttga tactcatact cttccttttt caatattatt gaagcattta
 11161 tcagggttat tgtctcatga gcggatacat atttgaatgt atttagaaaa ataaacaat
 11221 aggggttccg cgcacatttc cccgaaaagt gccac

30
35
40
45
50
55

SEQ ID NO:44 (pTnMod(Oval/ENT tag/Proins/PA) - QUAIL)

CTGACGCGCC CTGTAGCGGC GCATTAGCG CGCGGGTGT GGTGTTACG 50
 CGCAGCGTGA CCGCTACACT TGCCAGCGCC CTAGCGCCCG CTCCTTTCCG 100
 TTTCTTCCCT TCCTTTCTCG CCACGTTCCG CGGCATCAGA TTGGCTATTG 150
 GCCATTGCAT ACCTTGTATC CATATCATAA TATGTACATT TATATTGGCT 200
 CATGTCCAAC ATTACGCCCA TGTGACATT GATTATTGAC TAGTTATTAA 250
 TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA TGGAGTTCCG 300
 CGTTACATAA CTTACGGTAA ATGGCCCGCC TGGCTGACCG CCCAACGACC 350
 CCGCGCCATT GACGTCAATA ATGACGTATG TTCCCATAGT AACGCCAATA 400
 GGGACTTTCC ATTGACGTCA ATGGGTGGAG TATTTACGCT AAATGCCCCA 450
 CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTACGCC CCTATTGACG 500
 TCAATGACGG TAAATGGCCC GCCTGGCATT ATGCCAGTA CATGACCTTA 550
 TGGGACTTTC CTACTTGGCA GTACATCTAC GTATTAGTCA TCGCTATTAC 600
 CATGGTGTAT CGGTTTTTGGC AGTACATCAA TGGCGGTGGA TAGCGGTTTG 650
 ACTCACGGGG ATTTCCAAGT CTCCACCCCA TTGACGTCAA TGGGAGTTTG 700
 TTTTGGCACC AAAATCAACG GGACTTTCCA AAATGTCGTA ACAACTCCGC 750
 CCCATTGACG CAAATGGGCG GTAGCGGTGT ACGGTGGGAG GTCTATATAA 800
 GCAGAGCTCG TTAGTGAAC CGTCAGATCG CTTGGAGACG CCATCCACGC 850
 TGTTTTGACC TCCATAGAAG ACACCGGGAC CGATCCAGCC TCCGCGGCCG 900
 GGAACGGTGC ATTGGAACGC GGATTCCCCG TGCCAAGAGT GACGTAAGTA 950
 CCGCTATAG ACTCTATAG CACACCCCTT TGGCTCTTAT GCATGCTATA 1000
 CTGTTTTTGG CTGCGGCCCT ATACACCCCT GCTTCTTAT GCTATAGGTG 1050
 ATGGTATAGC TTAGCCTATA GGTGTGGGTT ATTGACCAAT ATTGACCACT 1100
 CCCCTATTGG TGACGATACT TTCCATTACT AATCCATAAC ATGGCTCTTT 1150
 GCCACAATA TCTCTATTGG CTATATGCCA ATACTCTGTC CTTGAGAGAC 1200
 TGACACGGAC TCTGTATTTT TACAGGATGG GGTCCCATTT ATTATTTACA 1250
 AATTCACATA TACAACAACG CCGTCCCCCG TGCCCGCAGT TTTTATTAAA 1300
 CATACCGTGG GATCTCCACG CGAATCTCGG GTACGTGTTT CGGACATGGG 1350
 CTCTTCTCCG GTAGCGGCGG AGCTTCCACA TCCGAGCCCT GGTCCCATGC 1400
 CTCCAGCGGC TCATGGTCCG TCGGCAGCTC CTGCTCTTA ACAGTGGAGG 1450
 CCAGACTTAG GCACAGCACA ATGCCACCA CCACAGTGT GCCGCACAAG 1500

	GCCGTGGCGG	TAGGGTATGT	GTCTGAAAAT	GAGCGTGGAG	ATTGGGCTCG	1550
	CACGGCTGAC	GCAGATGGAA	GACTTAAGGC	AGCGGCAGAA	GAAGATGCAG	1600
	GCAGCTGAGT	TGTTGTATTC	TGATAAGAGT	CAGAGGTAAC	TCCCGTTGCG	1650
5	GTGCTGTTAA	CGGTGGAGGG	CAGTGTAGTC	TGAGCAGTAC	TCGTTGCTGC	1700
	CGCGCGCGCC	ACCAGACATA	ATAGCTGACA	GACTAACAGA	CTGTTCCCTT	1750
	CCATGGGTCT	TTTCTGCAGT	CACCGTCGGA	CCATGTGTGA	ACTTGATATT	1800
	TTACATGATT	CTCTTTACCA	ATTCTGCCCC	GAATTACACT	TAAAAACGACT	1850
	CAACAGCTTA	ACGTTGGCTT	GCCACGCATT	ACTTGACTGT	AAAACTCTCA	1900
	CTCTTACCGA	ACTTGGCCGT	AACCTGCCAA	CCAAAGCGAG	AACAAAACAT	1950
10	AACATCAAAC	GAATCGACCG	ATTGTTAGGT	AATCGTCACC	TCCACAAAGA	2000
	GCGACTCGCT	GTATACCGTT	GGCATGCTAG	CTTTATCTGT	TCGGGAATAC	2050
	GATGCCCAT	GTACTTGTG	ACTGGTCTGA	TATTCGTGAG	CAAAAACGAC	2100
	TTATGGTATT	GCGAGCTTCA	GTCGCACTAC	ACGGTCGTTT	TGTTACTCTT	2150
	TATGAGAAAG	CGTTCGCGCT	TTCAGAGCAA	TGTTCAAAGA	AAGCTCATGA	2200
	CCAATTTCTA	GCCGACCTTG	CGAGCATTCT	ACCGAGTAAC	ACCACACCGC	2250
15	TCATTGTGAG	TGATGCTGGC	TTTAAAGTGC	CATGGTATAA	ATCCGTTGAG	2300
	AAGCTGGGTT	GGTACTGGTT	AAGTCGAGTA	AGAGGAAAAG	TACAATATGC	2350
	AGACCTAGGA	GCGGAAAAC	GGAAACCTAT	CAGCAACTTA	CATGATATGT	2400
	CATCTAGTCA	CTCAAAGACT	TTAGGCTATA	AGAGGCTGAC	TAAAAGCAAT	2450
	CCAATCTCAT	GCCAAATTCT	ATTGTATAAA	TCTCGCTCTA	AAGGCCGAAA	2500
20	AAATCAGCGC	TCGACACGGA	CTCATGTGCA	CCACCCGTC	CCTAAAATCT	2550
	ACTCAGCGTC	GGCAAAGGAG	CCATGGGTTT	TAGCAACTAA	CTTACCTGTT	2600
	GAAATTCGAA	CACCCAAACA	ACTTGTTAAT	ATCTATTGGA	AGCGAATGCA	2650
	GATTGAAGAA	ACCTTCCGAG	ACTTGAAAAG	TCCTGCCTAC	GGACTAGGCC	2700
	TACGCCATAG	CGGAACGAGC	AGCTCAGAGC	GTTTTGATAT	CATGCTGCTA	2750
	ATCGCCCTGA	TGCTTCAACT	AACATGTTGG	CTTGCGGGCG	TTCATGCTCA	2800
25	GAAACAAGGT	TGGGACAAGC	ACTTCCAGGC	TAACACAGTC	AGAAATCGAA	2850
	ACGTACTCTC	AACAGTTCGC	TTAGGCATGG	AAGTTTTGCG	GCATTCTGGC	2900
	TACACAATAA	CAAGGGAAGA	CTTACTCGTG	GCTGCAACCC	TACTAGCTCA	2950
	AAATTTATTC	ACACATGGTT	ACGCTTTGGG	GAAATTATGA	TAATGATCCA	3000
	GATCACTTCT	GGCTAATAAA	AGATCAGAGC	TCTAGAGATC	TGTTGTGTTG	3050
30	TTTTTTGTGG	ATCTGCTGTG	CCTTCTAGTT	GCCAGCCATC	TGTTGTTTGC	3100
	CCCTCCCCCG	TGCTTCTCTT	GACCTTGGA	GGTGCCACTC	CCACTGTCTT	3150
	TTCTCTAATA	AATGAGGAAA	TTGCATCGCA	TTGTCTGAGT	AGGTGTCATT	3200
	CTATTCTGGG	GGGTGGGGTG	GGGCAGCACA	GCAAGGGGGA	GGATTGGGAA	3250
	GACAATAGCA	GGCATGCTGG	GGATGCGGTG	GGCTCTATGG	GTACCTCTCT	3300
	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CTCTCGGTAC	CTCTCTCTCT	3350
35	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CGGTACCAGG	TGCTGAAGAA	3400
	TTGACCCGGT	GACCAAAGGT	GCCTTTTATC	ATCACTTTAA	AAATAAAAAA	3450
	CAATTACTCA	GTGCTGTGTA	TAAGCAGCAA	TTAATTATGA	TTGATGCCTA	3500
	CATCACAACA	AAAACGTGAT	TAACAAATGG	TTGGTCTGCC	TTAGAAAGTA	3550
	TATTTGAACA	TTATCTTGAT	TATATTATTG	ATAATAATAA	AAACCTTATC	3600
40	CCTATCCAAG	AAGTGATGCC	TATCATTTGG	TGGAATGAAC	TTGAAAAAAA	3650
	TTAGCCTTGA	ATACATTACT	GGTAAGGTAA	ACGCCATTGT	CAGCAAATTG	3700
	ATCCAAGAGA	ACCAACTTAA	AGCTTTCCTG	ACGGAATGTT	AATTCTCGTT	3750
	GACCCGTGAG	ACTGATGAAT	CCCTTAATGA	TTTTGGTAAA	AATCATTAA	3800
	TTAAGGTGGA	TACACATCTT	GTCAATATGAT	CCCGGTAATG	TGAGTTAGCT	3850
	CACTCATTAG	GCACCCAGG	CTTTACACTT	TATGCTTCCG	GCTCGTATGT	3900
45	TGTGTGGAAT	TGTGAGCGGA	TAACAATTTT	ACACAGGAAA	CAGCTATGAC	3950
	CATGATTACG	CCAAGCGCGC	AATTAACCCT	CACTAAAGGG	AACAAAAGCT	4000
	GGAGCTCCAC	CGCGGTGGCG	GCCGCTCTAG	AACTAGTGGA	TCCCCCGGGG	4050
	AGGTGAGAAT	GGTTTCTTTA	CTGTTTGTCA	ATTCTATTAT	TTCAATACAG	4100
	AACAAAAGCT	TCTATAACTG	AAATATATTT	GCTATTGTAT	ATTATGATTG	4150
50	TCCCTCGAAC	CATGAACACT	CCTCCAGCTG	AATTTTCAAA	TTCTCTGTCT	4200
	ATCTGCCAGG	CTGGAAGATC	ATGGAAGATC	TCTGAGGAAC	ATTGCAAGTT	4250
	CATACCATAA	ACTCATTTGG	AATTGAGTAT	TATTTTGCTT	TGAATGGAGC	4300
	TATGTTTTCG	AGTTCCCTCA	GAAGAAAAGC	TTGTTATAAA	GCGTCTACAC	4350
	CCATCAAAG	ATATATTTAA	ATATTCCAAC	TACAGAAAAG	TTTTGTCTGC	4400
	TCCTCACTCT	GATCTCAGTT	GGTTTCTTCA	CGTACATGCT	TCTTTATTG	4450
55	CCTATTTTGT	CAAGAAAATA	ATAGGTCAAG	TCCTGTCTCT	ACTTATCTCC	4500
	TGCCTAGCAT	GGCTTAGATG	CACGTTGTAC	ATTCAAGAAG	GATCAATGA	4550

	AACAGACTTC	TGGTCTGTTA	CAACAACCAT	AGTAATAAAC	AGACTAACTA	4600
	ATAATTGCTA	ATTATGTTTT	CCATCTCTAA	GGTTCCCACA	TTTTTCTGTT	4650
	TTAAGATCCC	ATTATCTGGT	TGTAACGTAA	GCTCAATGGA	ACATGAACAG	4700
5	TATTTCTCAG	TCTTTTCTCC	AGCAATCCTG	ACGGATTAGA	AGAACTGGCA	4750
	GAAAACACTT	TGTTACCCAG	AATTAAAAAC	TAATATTTGC	TCTCCCTTCA	4800
	ATCCAAAATG	GACCTATTGA	AACTAAAATC	TGACCCAATC	CCATTAAATT	4850
	ATTTCTATGG	CGTCAAAGGT	CAAACTTTTG	AAGGGAACCT	GTGGGTGGGT	4900
	CCCAATTGAG	GCTATATATT	CCCCAGGGCT	CAGCCAGTGG	ATCCATGGGC	4950
	TCCATCGGTG	CAGCAAGCAT	GGAATTTTGT	TTTGATGTAT	TCAAGGAGCT	5000
10	CAAAGTCCAC	CATGCCAATG	ACAACATGCT	CTACTCCCCC	TTTGCCATCT	5050
	TGTCAACTCT	GGCCATGGTC	TTCTTAGGTG	CAAAAGACAG	CACCAGGACC	5100
	CAGATAAATA	AGGTGTGTCA	CTTTGATAAA	CTTCCAGGAT	TCCGAGACAG	5150
	TATTGAAGCT	CAGTGTGGCA	CATCTGTAAA	TGTTCACTCT	TCACCTAGAG	5200
	ACATACTCAA	CCAAATCACC	AAACAAAATG	ATGCTTATTC	GTTTCAGCCTT	5250
	GCCAGTAGAC	TTTATGCTCA	AGAGACATAC	ACAGTCGTGC	CGGAATACCT	5300
15	GCAATGTGTG	AAGGAACTGT	ATAGAGGAGG	CTTAGAATCC	GTCAACTTTC	5350
	AAACAGCTGC	AGATCAAGCC	AGAGGCCTCA	TCAATGCCTG	GGTAGAAAGT	5400
	CAGACAAAAG	GAATTATCAG	AAACATCCTT	CAGCCAAGCT	CCGTGGATTG	5450
	TCAAACGTGA	ATGGTCCTGG	TTAATGCCAT	TGCCTTCAAG	GGACTGTGGG	5500
	AGAAAGCATT	TAAGGCTGAA	GACACGCAAA	CAATACCTTT	CAGAGTGACT	5550
20	GAGCAAGAAA	GCAAACCTGT	GCAGATGATG	TACCAGATTG	GTTCAATTTAA	5600
	AGTGGCATCA	ATGGCTTCTG	AGAAAATGAA	GATCCTGGAG	CTTCCATTTC	5650
	CCAGTGAAC	AATGAGCATG	TTGGTGCTGT	TGCCTGATGA	TGTCTCAGGC	5700
	CTTGAGCAGC	TTGAGAGTAT	AATCAGCTTT	GAAAACTGA	CTGAATGGAC	5750
	CAGTTCTAGT	ATTATGGAAG	AGAGGAAGGT	CAAAGTGATC	TTACCTCGCA	5800
	TGAAGATGGA	GGAGAAATAC	AACCTCACAT	CTCTCTTAAT	GGCTATGGGA	5850
25	ATTACTGACC	TGTTGAGCTC	TTGAGCCAA	CTGTCTGGCA	TCTCCTCAGT	5900
	AGGGAGCCTG	AAGATATCTC	AAGCTGTCCA	TGCAGCACAT	GCAGAAATCA	5950
	ATGAAGCGGG	CAGAGATGTG	GTAGGCTCAG	CAGAGGCTGG	AGTGGATGCT	6000
	ACTGAAGAAT	TTAGGGCTGA	CCATCCATTG	CTCTTCTGTG	TCAAGCACAT	6050
	CGAAACCAAC	GCCATTCTCC	TCCTTGGCAG	ATGTGTTTCT	CCGCGGCCAG	6100
30	CAGATGACGC	ACCAGCAGAT	GACGCACCAG	CAGATGACGC	ACCAGCAGAT	6150
	GACGCACCAG	CAGATGACGC	ACCAGCAGAT	GACGCAACAA	CATGTATCCT	6200
	GAAAGGCTCT	TGTGGCTGGA	TGGGCTGCT	GGATGACGAT	GACAAATTTG	6250
	TGAACCAACA	CCTGTGCGGC	TCACACCTGG	TGGAAGCTCT	CTACCTAGTG	6300
	TGCGGGGAAC	GAGGCTTCTT	CTACACACCC	AAGACCCGCC	GGGAGGCAGA	6350
	GGACCTGCAG	GTGGGGCAGG	TGGAGCTGGG	CGGGGGCCCT	GGTGCAGGCA	6400
35	GCCTGCAGCC	CTTGGCCCTG	GAGGGGTCCC	TGCAGAAGCG	TGGCATTGTG	6450
	GAACAATGCT	GTACCAGCAT	CTGCTCCCTC	TACCAGCTGG	AGAACTACTG	6500
	CAACTAGGGC	GCCTGGATCC	AGATCACTTC	TGGCTAATAA	AAGATCAGAG	6550
	CTCTAGAGAT	CTGTGTGTTG	GTTTTTTGTG	GATCTGCTGT	GCCTTCTAGT	6600
	TGCCAGCCAT	CTGTTGTTTG	CCCCTCCCCC	GTGCCTTCCT	TGACCCTGGA	6650
40	AGGTGCCACT	CCCACTGTCC	TTTCCTAATA	AAATGAGGAA	ATTGCATCGC	6700
	ATTGTCTGAG	TAGGTGTCAT	TCTATTCTGG	GGGGTGGGGT	GGGGCAGCAC	6750
	AGCAAGGGGG	AGGATTGGGA	AGACAATAGC	AGGCATGCTG	GGGATGCGGT	6800
	GGGCTCTATG	GGTACCTCTC	TCTCTCTCTC	TCTCTCTCTC	TCTCTCTCTC	6850
	TCTCTCGGTA	CCTCTCTCGA	GGGGGGGCC	GGTACCCAAT	TGCCCCTATA	6900
	GTGAGTCGTA	TTACGCGCGC	TCACTGGCCG	TCGTTTACAA	ACGTCGTGAC	6950
45	TGGGAAAACC	CTGGCGTTAC	CCAACCTAAT	CGCCTTGCA	CACATCCCCC	7000
	TTTCGCCAGC	TGGCGTAATA	GCGAAGAGGC	CCGCACCGAT	CGCCCTTCCC	7050
	AACAGTTGCG	CAGCCTGAAT	GGCGAATGGA	AATTGTAAGC	GTTAATATTT	7100
	TGTTAAAATT	CGCGTTAAAT	TTTTGTAAAT	TCAGCTCATT	TTTAAACCAA	7150
	TAGGCCGAAA	TGGGCAAAAT	CCCTTATAAA	TCAAAAAGAT	AGACCGAGAT	7200
	AGGGTTGAGT	GTGTGTTCCAG	TTTGGAACAA	GAGTCCACTA	TTAAAGAACG	7250
50	TGGACTCCAA	CGTCAAAGGG	CGAAAAACCG	TCTATCAGGG	CGATGGCCCA	7300
	CTACTCCGGG	ATCATATGAC	AAGATGTGTA	TCCACCTTAA	CTTAATGATT	7350
	TTTACCAAAA	TCATTAGGGG	ATTATCAGT	GCTCAGGGTC	AACGAGAATT	7400
	AACATTCCGT	CAGGAAAGCT	TATGATGATG	ATGTGCTTAA	AAACTTACTC	7450
	AATGGCTGGT	TATGCATATC	GCAATACATG	CGAAAAACCT	AAAAGAGCTT	7500
55	GCCGATAAAA	AAGGCCAATT	TATTGCTATT	TACCGCGGCT	TTTTATTGAG	7550
	CTTGAAAGAT	AAATAAAATA	GATAGGTTTT	ATTTGAAGCT	AAATCTTCTT	7600

	TATCGTAAAA	AATGCCCTCT	TGGGTATCA	AGAGGGTCAT	TATATTTTCGC	7650
	GGAATAACAT	CATTTGGTGA	CGAAATAACT	AAGCACTTGT	CTCCTGTTTA	7700
5	CTCCCCGAG	CTTGAGGGGT	TAACATGAAG	GTCATCGATA	GCAGGATAAT	7750
	AATACAGTAA	AACGCTAAAC	CAATAATCCA	AATCCAGCCA	TCCCAAATTG	7800
	GTAGTGAATG	ATTATAAATA	ACAGCAAACA	GTAATGGGCC	AATAACACCG	7850
	GTTGCATTGG	TAAGGCTCAC	CAATAATCCC	TGTAAAGCAC	CTTGCTGATG	7900
	ACTCTTTGTT	TGGATAGACA	TCACTCCCTG	TAATGCAGGT	AAAGCGATCC	7950
	CACCACCAGC	CAATAAAATT	AAAACAGGGA	AACTAACC	ACCTTCAGAT	8000
10	ATAAACGCTA	AAAAGGCCAA	TGCACTACTA	TCTGCAATAA	ATCCGAGCAG	8050
	TACTGCCGTT	TTTTCGCCCC	ATTTAGTGGC	TATTCTTCCT	GCCACAAAGG	8100
	CTTGGAATAC	TGAGTGTA	AGACCAAGAC	CCGCTAATGA	AAAGCCAACC	8150
	ATCATGCTAT	TCCATCCAAA	ACGATTTTCG	GTAATAGCA	CCCACACCGT	8200
	TGCGGGGAAT	TGGCCTATCA	ATTGCGCTGA	AAAATAAATA	ATCAACAAAA	8250
	TGGCATCGTT	TTAAATAAAG	TGATGTATAC	CGAATTCAGC	TTTTGTTCCT	8300
15	TTTAGTGAGG	GTTAATTGCG	CGCTTGCGGT	AATCATGGTC	ATAGCTGTTT	8350
	CCTGTGTGAA	ATTGTTATCC	GCTCACAATT	CCACACAACA	TACGAGCCGG	8400
	AAGCATAAAG	TGTAAAGCCT	GGGGTGCCTA	ATGAGTGAGC	TAACCTACAT	8450
	TAATTGCGTT	GCGCTCACTG	CCCGCTTCC	AGTCGGGAAA	CCTGTCGTGC	8500
	CAGCTGCATT	AATGAATCGG	CCAACGCGCG	GGGAGAGGCG	GTTTGCCTAT	8550
20	TGGGCGCTCT	TCCGCTTCCT	CGCTCACTGA	CTCGCTGCGC	TCCGTCGTTT	8600
	GGCTGCGGCG	AGCGGTATCA	GCTCACTCAA	AGGCGGTAAT	ACGGTTATCC	8650
	ACAGAATCAG	GGGATAACGC	AGGAAAGAAC	ATGTGAGCAA	AAGGCCAGCA	8700
	AAAGGCCAGG	AACCGTAAAA	AGGCCGCGTT	GCTGGCGTTT	TTCCATAGGC	8750
	TCCGCCCCCC	TGACGAGCAT	CACAAAATC	GACGCTCAAG	TCAGAGGTGG	8800
	CGAAACCCGA	CAGGACTATA	AAGATACCAG	GCGTTTCCCC	CTGGAAGCTC	8850
25	CCTCGTGCGC	TCTCCTGTTT	CGACCCTGCC	GCTTACCGGA	TACCTGTCCG	8900
	CCTTTCTCCC	TTCCGGAAGC	GTGGCGCTTT	CTCATAGCTC	ACGCTGTAGG	8950
	TATCTCAGTT	CGGTGTAGGT	CGTTCGCTCC	AAGCTGGGCT	GTGTGCACGA	9000
	ACCCCCCGTT	CAGCCCGACC	GCTGCGCCTT	ATCCGGTAAC	TATCGTCTTG	9050
	AGTCCAACCC	GGTAAGACAC	GACTTATCGC	CACTGGCAGC	AGCCACTGGT	9100
30	AACAGGATTA	GCAGAGCGAG	GTATGTAGGC	GGTGCTACAG	AGTCTTGAA	9150
	GTGGTGGCCT	AACCTACGGT	ACACTAGAAG	GACAGTATTT	GGTATCTGCG	9200
	CTCTGCTGAA	GCCAGTTACC	TTCCGAAAAA	GAGTTGGTAG	CTCTTGATCC	9250
	GGCAAACAAA	CCACCGCTGG	TAGCGGTGGT	TTTTTTGTTT	GCAAGCAGCA	9300
	GATTACGCGC	AGAAAAAAG	GATCTCAAGA	AGATCCTTTG	ATCTTTTCTA	9350
	CGGGGTCTGA	CGCTCAGTGG	AACGAAAAT	CACGTTAAGG	GATTTTGGTC	9400
35	ATGAGATTAT	CAAAAAGGAT	CTTCACCTAG	ATCCTTTTAA	ATTAAAAAATG	9450
	AAGTTTTTAA	TCAATCTAAA	GTATATATGA	GTAAACTTGG	TCTGACAGTT	9500
	ACCAATGCTT	AATCAGTGAG	GCACCTATCT	CAGCGATCTG	TCTATTTCTG	9550
	TCATCCATAG	TTGCCTGACT	CCCCGTGCTG	TAGATAACTA	CGATACGGGA	9600
	GGGCTTACCA	TCTGGCCCCA	GTGCTGCAAT	GATACCGCGA	GACCCACGCT	9650
40	CACCGGCTCC	AGATTTATCA	GCAATAAACC	AGCCAGCCGG	AAGGGCCGAG	9700
	CGCAGAGTGG	GTCCGTGCAAC	TTTATCCGCC	TCCATCCAGT	CTATTAAATTG	9750
	TTGCCGGGAA	GCTAGAGTAA	GTAGTTCGCC	AGTTAATAGT	TTGCCGCAACG	9800
	TTGTTGCCAT	TGCTACAGGC	ATCGTGTGT	CACGCTCGTC	GTTTGGTATG	9850
	GCTTCATTCA	GCTCCGGTTC	CCAACGATCA	AGGCGAGTTA	CATGATCCCC	9900
	CATGTTGTGC	AAAAAAGCGG	TTAGCTCCTT	CGGTCTCTCC	ATCGTTGTCA	9950
45	GAAGTAAGTT	GGCCGAGTGG	TTATCACTCA	TGGTTATGGC	AGCACTGCAT	10000
	AATTCTCTTA	CTGTCAATGCC	ATCCGTAAGA	TGCTTTTCTG	TGACTGGTGA	10050
	GTAATCAACC	AAGTCATTCT	GAGAATAGTG	TATGCGGCGA	CCGAGTTGCT	10100
	CTTGCCCGGC	GTCAATACGG	GATAATACCG	CGCCACATAG	CAGAACTTTA	10150
	AAAGTGCTCA	TCATTGGA	ACGTTCTTCG	GGGCGAAAAC	TCTCAAGGAT	10200
	CTTACCGCTG	TTGAGATCCA	GTTGATGTA	ACCCACTCGT	GCACCCAACT	10250
50	GATCTTCAGC	ATCTTTTACT	TTCACCAAGC	TTTCTGGGTG	AGCAAAAACA	10300
	GGAAGGCAAA	ATGCCGCAAA	AAAGGGAATA	AGGGCGACAC	GGAAATGTTG	10350
	AATACTCATA	CTCTTCCTTT	TTCAATATTA	TTGAAGCATT	TATCAGGGTT	10400
	ATTGTCTCAT	GAGCGGATAC	ATATTTGAAT	GTATTTAGAA	AAATAAACAA	10450
55	ATAGGGGTTT	CGCGCACATT	TCCCCGAAAA	GTGCCAC		10487

SEQ ID NO:45 (pTnMod(Oval/ENT tag/P146/PA) - Chicken)

	CTGACGCGCC	CTGTAGCGGC	GCATTAAGCG	CGGCGGGTGT	GGTGGTTACG	50
	CGCAGCGTGA	CCGCTACACT	TGCCAGCGCC	CTAGCGCCCG	CTCCTTTCCG	100
5	TTTCTTCCCT	TCCTTTCTCG	CCACGTTCCG	CGGCATCAGA	TTGGCTATTG	150
	GCCATTGCAT	ACGTTGTATC	CATATCATAA	TATGTACATT	TATATTGGCT	200
	CATGTCCAAC	ATTACCGCCA	TGTTGACATT	GATTATTGAC	TAGTTATTAA	250
	TAGTAATCAA	TTACGGGGTC	ATTAGTTCAT	AGCCCATATA	TGGAGTTCCG	300
	CGTTACATAA	CTTACGGTAA	ATGGCCCCGC	TGGCTGACCG	CCCAACGACC	350
	CCGCCCCATT	GACGTCAATA	ATGACGTATG	TTCCCATAGT	AACGCCAATA	400
10	GGGACTTTCC	ATTGACGTCA	ATGGGTGGAG	TATTTACGGT	AAACTGCCCA	450
	CTTGGCAGTA	CATCAAGTGT	ATCATATGCC	AAGTACGCCC	CCTATTGACG	500
	TCAATGACGG	TAAATGGCCC	GCCTGGCATT	ATGCCCAGTA	CATGACCTTA	550
	TGGGACTTTC	CTACTTGGCA	GTACATCTAC	GTATTAGTCA	TCGCTATTAC	600
	CATGGTGATG	CGGTTTGGC	AGTACATCAA	TGGGCGTGGA	TAGCGGTTTG	650
	ACTCACGGGG	ATTTCCAAGT	CTCCACCCCA	TTGACGTCAA	TGGGAGTTTG	700
15	TTTTGGCACC	AAAATCAACG	GGACTTTCCA	AAATGTCGTA	ACAACCTCCG	750
	CCCATTTGACG	CAAATGGGCG	GTAGGCGTGT	ACGGTGGGAG	GTCTATATAA	800
	GCAGAGCTCG	TTTAGTGAAC	CGTCAGATCG	CCTGGAGACG	CCATCCACGC	850
	TGTTTTGACC	TCCATAGAAG	ACACCGGGAC	CGATCCAGCC	TCCGCGGCCG	900
	GGAACGGTGC	ATTGGAACGC	GGATTCCCCG	TGCCAAGAGT	GACGTAAGTA	950
20	CCGCTATAG	ACTCTATAGG	CACACCCTT	TGGCTCTTAT	GCATGCTATA	1000
	CTGTTTTTGG	CTTGGGGCCT	ATACACCCCC	GCTTCCTTAT	GCTATAGGTG	1050
	ATGGTATAGC	TTAGCCTATA	GGTGTGGGTT	ATTGACCATT	ATTGACCACT	1100
	CCCCTATTGG	TGACGATACT	TTCCATTACT	AATCCATAAC	ATGGCTCTTT	1150
	GCCACAACCTA	TCTCTATTGG	CTATATGCCA	ATACTCTGTC	CTTCAGAGAC	1200
	TGACACGGAC	TCTGTATTTT	TACAGGATGG	GGTCCCATTT	ATTATTTACA	1250
25	AATTACATA	TACAACAACG	CCGTCCCCCG	TGCCCGCAGT	TTTTATTAAA	1300
	CATAGCGTGG	GATCTCCACG	CGAATCTCGG	GTACGTGTTT	CGGACATGGG	1350
	CTCTTCTCCG	GTAGCGGCGG	AGCTTCCACA	TCCGAGCCCT	GGTCCCATGC	1400
	CTCCAGCGGC	TCATGGTCCG	TCGGCAGCTC	CTTGCTCCTA	ACAGTGGAGG	1450
	CCAGACTTAG	GCACAGCACA	ATGCCACCA	CCACCAGTGT	GCCGCACAAG	1500
30	GCGGTGGCGG	TAGGGTATGT	GTCTGAAAT	GAGCGTGGAG	ATTGGGCTCG	1550
	CACGGCTGAC	GCAGATGGAA	GACTTAAGGC	AGCGGCAGAA	GAAGATGCAG	1600
	GCAGCTGAGT	TGTTGTATTC	TGATAAGAGT	CAGAGGTAAC	TCCCGTTGCG	1650
	GTGCTGTTAA	CGGTGGAGGG	CAGTGTAGTC	TGAGCAGTAC	TCGTGCTGCG	1700
	CGCGCGCGCC	ACCAGACATA	ATAGCTGACA	GACTAACAGA	CTGTTCTCTT	1750
	CCATGGGTCT	TTTCTGCAGT	CACCGTCGGA	CCATGTGTGA	ACTTGATATT	1800
35	TTACATGATT	CTCTTTACCA	ATTCTGCCCC	GAATTACACT	TAAAACGACT	1850
	CAACAGCTTA	ACGTTGGCTT	GCCACGCATT	ACTTGACTGT	AAAACCTCTA	1900
	CTCTTACCGA	ACTTGGCCGT	AACCTGCCAA	CCAAAGCGAG	AACAAAACAT	1950
	AACATCAAAC	GAATCGACCG	ATTGTTAGGT	AATCGTCACC	TCCACAAAGA	2000
	GCGACTCGCT	GTATACCGTT	GGCATGCTAG	CTTTATCTGT	TCGGGAATAC	2050
	GATGCCCATT	GTACTTGTTG	ACTGGTCTGA	TATTCGTGAG	CAAAAACGAC	2100
40	TTATGGTATT	GCGAGCTTCA	GTCGCACTAC	ACGGTCTGTT	TGTTACTCTT	2150
	TATGAGAAAG	CGTTCCCGCT	TTCAGAGCAA	TGTTCAAAGA	AAGCTCATGA	2200
	CCAATTTCTA	GCCGACCTTG	CGAGCATTCT	ACCGAGTAAC	ACCACACCGC	2250
	TCATTGTCTAG	TGATGCTGGC	TTTAAAGTGC	CATGGTATAA	ATCCGTTGAG	2300
	AAGCTGGGTT	GGTACTGGTT	AAGTCGAGTA	AGAGGAAAAG	TACAATATGC	2350
45	AGACCTAGGA	GCGGAAAAC	GGAAACCTAT	CAGCAACTTA	CATGATATGT	2400
	CATCTAGTCA	CTCAAAGACT	TTAGGCTATA	AGAGGCTGAC	TAAAAGCAAT	2450
	CCAATCTCAT	GCCAAATTCT	ATTGTATAAA	TCTCGCTCTA	AAGGCCGAAA	2500
	AAATCAGCGC	TCGACACGGA	CTCATTTGTC	CCACCCGTCA	CCTAAAATCT	2550
	ACTCAGCGTC	GGCAAAGGAG	CCATGGGTTT	TAGCAACTAA	CTTACCTGTT	2600
	GAAATTCGAA	CACCCAAACA	ACTTGTTAAT	ATCTATTGCA	AGCGAATGCA	2650
50	GATTGAAGAA	ACCTTCGAG	ACTTGAAAAG	TCCTGCCTAC	GGACTAGGCC	2700
	TACGCCATAG	CCGAACGAGC	AGCTCAGAGC	GTTTTGATAT	CATGCTGCTA	2750
	ATCGCCCTGA	TGCTTCAACT	AACATGTTGG	CTTGCGGGCG	TTCATGCTCA	2800
	GAAACAAAGT	TGGGACAAAG	ACTTCCAGGC	TAACACAGTC	AGAAATCGAA	2850
	ACGTACTCTC	AACAGTTCCG	TTAGGCATGG	AAGTTTTCGG	GCATTCTGGC	2900
55	TACACAATAA	CAAGGGAAGA	CTTACTCGTG	GCTGCAACCC	TACTAGCTCA	2950
	AAATTTATTC	ACACATGGTT	ACGCTTTGGG	GAAATTATGA	TAATGATCCA	3000
	GATCACTTCT	GGCTAATAAA	AGATCAGAGC	TCTAGAGATC	TGTGTGTTGG	3050

	TTTTTTGTGG	ATCTGCTGTG	CCTTCTAGTT	GCCAGCCATC	TGTTGTTTGC	3100
	CCCTCCCCCG	TGCTTCTCTT	GACCCCTGGAA	GGTGCCACTC	CCACTGTCTT	3150
	TTCTTAATAA	AATGAGGAAA	TTGCATCGCA	TTGTCTGAGT	AGGTGTCTAT	3200
5	CTATTCTGGG	GGGTGGGGTG	GGGCAGCACA	GCAAGGGGGA	GGATTGGGAA	3250
	GACAATAGCA	GGCATGCTGG	GGATGCGGTG	GGCTCTATGG	GTACCTCTCT	3300
	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CTCTCGGTAC	CTCTCTCTCT	3350
	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CGGTACCAGG	TGCTGAAGAA	3400
	TTGACCCGGT	GACCAAAGGT	GCCTTTTATC	ATCACTTTAA	AAATAAAAAA	3450
	CAATTACTCA	GTGCCTGTTA	TAAGCAGCAA	TTAATTATGA	TTGATGCCTA	3500
10	CATCACAAACA	AAAACCTGAT	TAACAAATGG	TTGGTCTGCC	TTAGAAAGTA	3550
	TATTTGAACA	TTATCTTGAT	TATATTATTG	ATAATAATAA	AAACCTTATC	3600
	CCTATCCAAG	AAGTGATGCC	TATCATTTGGT	TGGAATGAAC	TTGAAAAAAA	3650
	TTAGCCTTGA	ATACATTACT	GGTAAGGTAA	ACGCCATTGT	CAGCAAATTG	3700
	ATCCAAGAGA	ACCAACTTAA	AGCTTTCCTG	ACGGAATGTT	AATTCTCGTT	3750
15	GACCCCTGAG	ACTGATGAAT	CCCCTAATGA	TTTTGGTAAA	AATCATTAA	3800
	TTAAGGTGGA	TACACATCTT	GTCATATGAT	CCCGGTAATG	TGAGTTAGCT	3850
	CACTCATTAG	GCACCCAGG	CTTTACACTT	TATGCTTCCG	GCTCGTATGT	3900
	TGTGTGGAAT	TGTGAGCGGA	TAACAATTTT	ACACAGGAAA	CAGCTATGAC	3950
	CATGATTACG	CCAAGCGCGC	AATTAACCTT	CACTAAAGGG	AACAAAAGCT	4000
20	GGAGCTCCAC	CGCGGTGGCG	GCCGCTCTAG	AACTAGTGGA	TCCCCGGGG	4050
	AGGTGGAAT	GGTTTCTTTA	CTGTTTGTCA	ATTCTATTAT	TTCAATACAG	4100
	AACAATAGCT	TCTATAACTG	AAATATATTT	GCTATTGTAT	ATTATGATTG	4150
	TCCTTCGAAC	CATGAACACT	CCTCCAGCTG	AATTTTACAA	TTCTCTGTCT	4200
	ATCTGCCAGG	CCATTAAAGT	ATTCATGGAA	GATCTTTGAG	GAACACTGCA	4250
	AGTTTCATATC	ATAAACACAT	TTGAAATTGA	GTATTGTTTT	GCATGTATG	4300
25	GAGCTATGTT	TTGCTGTATC	CTCAGAAAAA	AAGTTTGTTA	TAAAGCATT	4350
	ACACCCATAA	AAAGATAGAT	TTAAATATTC	CAGCTATAGG	AAAGAAAGTG	4400
	CGTCTGCTCT	TCACCTTAGT	CTCAGTTGGC	TCCTTCACAT	GCATGCTTCT	4450
	TTATTTCTCC	TATTTTGTCA	AGAAAATAAT	AGGTACAGTC	TTGTTCTCAC	4500
	TTATGTCCTG	CCTAGCATGG	CTCAGATGCA	CGTTGTAGAT	ACAAGAAGGA	4550
30	TCAAATGAAA	CAGACTTCTG	GTCTGTTACT	ACAACCATAG	TAATAAGCAC	4600
	ACTAATAAT	AATTGCTAAT	TATGTTTTTC	ATCTCTAAGG	TTCCACATT	4650
	TTTCTGTTTT	CTTAAAGATC	CCATTATCTG	GTTGTAACCTG	AAGTCAATG	4700
	GAACATGAGC	AATATTTCCT	AGTCTTCTCT	CCCATCCAAC	AGTCCTGATG	4750
	GATTAGCAGA	ACAGGCAGAA	AACACATTGT	TACCCAGAAT	TAAAACTAA	4800
	TATTTGCTCT	CCATTCAATC	CAAAATGGAC	CTATTGAAAC	TAAAATCTAA	4850
35	CCCAATCCCA	TTAAATGATT	TCTATGGCGT	CAAAGGTCAA	ACTTCTGAAG	4900
	GGAACCTGTG	GGTGGGTCAC	AATTCAGGCT	ATATATTCCC	CAGGGCTCAG	4950
	CGGATCCATG	GGCTCCATCG	GCGCAGCAAG	CATGGAATTT	TGTTTGTATG	5000
	TATTTCAAGGA	GCTCAAAGTC	CACCATGCCA	ATGAGAACAT	CTTCTACTGC	5050
	CCCATTTGCCA	TCATGTCAGC	TCTAGCCATG	GTATACCTGG	GTGCAAAAGA	5100
	CAGCACCAGG	ACACAGATAA	ATAAGGTTGT	TCGCTTTGAT	AAACTTCCAG	5150
40	GATTCGGAGA	CAGTATTGAA	GCTCAGTGTG	GCACATCTGT	AAACGTTTCC	5200
	TCTTCACTTA	GAGACATCCT	CAACCAATC	ACCAAACCAA	ATGATGTTTA	5250
	TTCTGTTGAG	CTTGCCAGTA	GACTTTATGC	TGAAGAGAGA	TACCCAATCC	5300
	TGCCAGAATA	CTTGCACTGT	GTGAAGGAAC	TGTATAGAGG	AGGCTTGGAA	5350
	CCTATCAACT	TTCAAACAGC	TGCAGATCAA	GCCAGAGAGC	TCATCAATTC	5400
45	CTGGGTAGAA	AGTCAGACAA	ATGGAATTAT	CAGAAATGTC	CTTCAGCCAA	5450
	GCTCCGTGGA	TTCTCAAAC	GCAATGGTTC	TGGTTAATGC	CATTGTCTTC	5500
	AAAGGACTGT	GGGAGAAAAC	ATTTAAGGAT	GAAGACACAC	AAGCAATGCC	5550
	TTTCAGAGTG	ACTGAGCAAG	AAAGCAAACC	TGTGCAGATG	ATGTACCAGA	5600
	TTGGTTTATT	TAGAGTGGCA	TCAATGGCTT	CTGAGAAAAT	GAAGATCCTG	5650
	GAGCTTCCAT	TTGCCAGTGG	GACAATGAGC	ATGTTGGTGC	TGTTGCCTGA	5700
50	TGAAGTCTCA	GGCCTTGAGC	AGCTTGAGAG	TATAATCAAC	TTTGAAAAAC	5750
	TGACTGAATG	GACCACTTCT	AATGTTATGG	AAGAGAGGAA	GATCAAAGTG	5800
	TACTTACCTC	GCATGAAGAT	GGAGGAAAAA	TACAACCTCA	CATCTGTCTT	5850
	AATGGCTATG	GGCATTACTG	ACGTGTTTAG	CTCTTCAGCC	AATCTGTCTG	5900
	GCATCTCCTC	AGCAGAGAGC	CTGAAGATAT	CTCAAGCTGT	CCATGCAGCA	5950
	CATGCAGAAA	TCAATGAAGC	AGGCAGAGAG	GTGGTAGGGT	CAGCAGAGGC	6000
55	TGGAGTGGAT	GCTGCAAGCG	TCTCTGAAGA	ATTTAGGGCT	GACCATCCAT	6050
	TCCTCTTCTG	TATCAAGCAC	ATCGCAACCA	ACGCCGTTCT	CTTCTTTGGC	6100

	AGATGTGTTT	CCCCTCCGCG	GCCAGCAGAT	GACGCACCAG	CAGATGACGC	6150
	ACCAGCAGAT	GACGCACCAG	CAGATGACGC	ACCAGCAGAT	GACGCACCAG	6200
5	CAGATGACGC	AACAACATGT	ATCCTGAAAG	GCTCTTGTTG	CTGGATCGGC	6250
	CTGCTGGATG	ACGATGACAA	AAAATACAAA	AAAGCACTGA	AAAACTGGC	6300
	AAAACTGCTG	TAATGAGGGC	GCCTGGATCC	AGATCACTTC	TGGCTAATAA	6350
	AAGATCAGAG	CTCTAGAGAT	CTGTGTGTTG	GTITTTTGTG	GATCTGCTGT	6400
	GCCTTCTAGT	TGCCAGCCAT	CTGTTGTTTG	CCCCTCCCCC	GTGCCTTCCT	6450
	TGACCCTGGA	AGGTGCCACT	CCCACTGTCC	TTTCTAATA	AAATGAGGAA	6500
10	ATTGCATCGC	ATTGCTGAG	TAGGTGTCAT	TCTATTCTGG	GGGGTGGGGT	6550
	GGGGCAGCAC	AGCAAGGGGG	AGGATTGGGA	AGACAATAGC	AGGCATGCTG	6600
	GGGATGCGGT	GGGCTCTATG	GGTACCTCTC	TCTCTCTCTC	TCTCTCTCTC	6650
	TCTCTCTCTC	TCTCTCGTA	CCTCTCTCGA	GGGGGGGGCC	GGTACCCAAT	6700
	TCGCCCTATA	GTGAGTCGTA	TTACGCGCGC	TCACTGGCCG	TCGTTTACAA	6750
	ACGTCGTGAC	TGGGAAAACC	CTGGCGTTAC	CCAACCTAAT	CGCCTTGACG	6800
15	CACATCCCCC	TTTCGCCAGC	TGGCGTAATA	GCGAAGAGGC	CCGCACCGAT	6850
	CGCCCTTCCC	AACAGTTGCG	CAGCCTGAAT	GGCGAATGGA	AATTGTAAGC	6900
	GTTAATATTT	TGTTAAAATT	CGCGTTAAAT	TTTTGTATAA	TCAGCTCATT	6950
	TTTTAAACAA	TAGGCCGAAA	TCGGCAAAAT	CCCTTATAAA	TCAAAAGAAT	7000
	AGACCAGAGAT	AGGGTTGAGT	GTGTTCCAG	TTTGGAACAA	GAGTCCACTA	7050
20	TTAAGAAACG	TGGACTCCAA	CGTCAAAGGG	CGAAAAACCG	TCTATCAGGG	7100
	CGATGGCCCA	CTACTCCGGG	ATCATATGAC	AAGATGTGTA	TCCACCTTAA	7150
	CTTAATGATT	TTTACCAGAA	TCATTAGGGG	ATTCACTAGT	GCTCAGGGTC	7200
	AACGAGAATT	AACATTCCGT	CAGGAAAGCT	TATGATGATG	ATGTGCTTAA	7250
	AAACTTACTC	AATGGCTGGT	TATGCATATC	GCAATACATG	CGAAAAACCT	7300
	AAAAGAGCTT	GCCGATAAAA	AAGGCCAATT	TATTGCTATT	TACCGCGGCT	7350
25	TTTTATTGAG	CTTGAAAGAT	AAATAAAATA	GATAGGTTTT	ATTTGAAGCT	7400
	AAATCTTCTT	TATCGTAAAA	AATGCCCTCT	TGGGTTATCA	AGAGGGTCAT	7450
	TATATTTTCG	GGAATAACAT	CATTTGGTGA	CGAAATAACT	AAGCACTTGT	7500
	CTCCTGTTTA	CTCCCTGAG	CTTGAGGGGT	TAACATGAAG	GTCATCGATA	7550
	GCAGGATAAT	AATACGTAA	AACGCTAAAC	CAATAATCCA	AATCCAGCCA	7600
30	TCCCAAATYG	GTAGTGAATG	ATTATAAAATA	ACAGCAAAACA	GTAATGGGCC	7650
	AATAACACCG	GTTGCATTGG	TAAGGCTCAC	CAATAATCCC	TGTAAAGCAC	7700
	CTTGCTGATG	ACTCTTTGTT	TGGATAGACA	TCACTCCCTG	TAATGCAGGT	7750
	AAAGCGATCC	CACCACCAGC	CAATAAAATT	AAAACAGGGA	AAACTAACCA	7800
	ACCTTCAGAT	ATAAACCGTA	AAAAGGCCAA	TGCACTACTA	TCTGCAATAA	7850
	ATCCGAGCAG	TACTGCCGTT	TTTTCGCCCC	ATTTAGTGSC	TATTCTTCCT	7900
35	GCCACAAAGG	CTTGGAAATC	TGAGTGTAAG	AGACCAAGAC	CCGCTAATGA	7950
	AAAGCCAACC	ATCATGCTAT	TCCATCCAAA	ACGATTTTCG	GTAAATAGCA	8000
	CCCACACCGT	TGCGGGAATT	TGGCCTATCA	ATTGCGCTGA	AAAAATAATA	8050
	ATCAACAAAA	TGGCATCGTT	TTAAATAAAG	TGATGTATAC	CGAATTCAGC	8100
	TTTTGTTCCC	TTTAGTGAGG	GTAAATTGCG	CGCTTGGCGT	AATCATGGTC	8150
	ATAGCTGTTT	CCTGTGTGAA	ATGTATTATC	GCTCACAATT	CCACACAACA	8200
40	TACGAGCCGG	AAGCATAAAG	TGTAAAGCCT	GGGGTGCCTA	ATGAGTGAGC	8250
	TAACTCACAT	TAATTGCGTT	GCGCTCACTG	CCCGCTTTCC	AGTCGGGAAA	8300
	CCTGTCTGTC	CAGCTGCATT	AATGAATCGG	CCAACGCGCG	GGGAGAGGCG	8350
	GTTTGCGTAT	TGGGCGCTCT	TCCGCTTCCT	CGCTCACTGA	CTCGCTGCGC	8400
	TCGGTCGTTT	GGCTGCGGCG	AGCGGTATCA	GCTCACTCAA	AGGCGGTAAT	8450
45	ACGGTTATCC	ACAGAATCAG	GGGATAACGC	AGGAAAGAAC	ATGTGAGCAA	8500
	AAGGCCAGCA	AAAGGCCAGG	AACCGTAAAA	AGGCCGCGTT	GCTGGCGTTT	8550
	TTCCATAGGC	TCCGCCCCCC	TGACGAGCAT	CACAAAAATC	GACGCTCAAG	8600
	TCAGAGGTGG	CGAAACCCGA	CAGGACTATA	AAGATACCAG	GCGTTTCCCC	8650
	CTGGAAGCTC	CCTCGTGCGC	TCTCCTGTTT	CGACCCTGCC	GCTTACCGGA	8700
	TACCTGTCCG	CCTTCTCTCC	TTCCGGGAAG	GTGGCGCTTT	CTCATAGCTC	8750
50	ACGCTGTAGG	TATCTCAGTT	CGGTGTAGGT	CGTTCGCTCC	AAGCTGGGCT	8800
	GTGTGCACGA	ACCCCCGTTT	CAGCCCCACC	GCTGCGCCTT	ATCCGGTAAC	8850
	TATCCTCTTG	AGTCCAACCC	GGTAAGACAC	GACTTATCGC	CACTGGCAGC	8900
	AGCCACTGGT	AACAGGATTA	GCAGAGCGAG	GTATGTAGGC	GGTGCTACAG	8950
	AGTTCTTGAA	GTGGTGCCCT	AACTACGGCT	ACACTAGAAG	GACAGTATTT	9000
	GGTATCTGCG	CTCTGCTGAA	GCCAGTTACC	TTCCGAAAAA	GAGTTGGTAG	9050
55	CTCTTGATCC	GGCAACAAAA	CCACCGCTGG	TAGCGGTGGT	TTTTTTGTTT	9100
	GCAAGCAGCA	GATTACGCGC	AGAAAAAAG	GATCTCAAGA	AGATCCTTTG	9150

ATCTTTTCTA CGGGGTCTGA CGCTCAGTGG AACGAAAAC CACGTTAAGG 9200
 GATTTTGGTC ATGAGATTAT CAAAAGGAT CTTACCTAG ATCCTTTTAA 9250
 ATTA AAAATG AAGTTTTAAA TCAATCTAAA GTATATATGA GTAAACTTGG 9300
 TCTGACAGTT ACCAATGCTT AATCAGTGAG GCACCTATCT CAGCGATCTG 9350
 TCTATTTCGT TCATCCATAG TTGCCTGACT CCCCCTCGTG TAGATAACTA 9400
 CGATACGGGA GGGCTTACCA TCTGGCCCCA GTGCTGCAAT GATACCGCGA 9450
 GACCCACGCT CACCGGCTCC AGATTTATCA GCAATAAAC AGCCAGCCGG 9500
 AAGGGCCGAG CGCAGAAAGT GTCTGCAAC TTTATCCGCC TCCATCCAGT 9550
 CTATTAATTG TTGCCGGGAA GCTAGAGTAA GTAGTTCGCC AGTTAATAGT 9600
 TTGCGCAACG TTGTTGCCAT TGCTACAGGC ATCGTGGTGT CACGCTCGTC 9650
 GTTTGGTATG GCTTCATTCA GCTCCGGTTC CCAACGATCA AGGCGAGTTA 9700
 CATGATCCCC CATGTTGTGC AAAAAAGCGG TTAGCTCCTT CGGTCCTCCG 9750
 ATCGTTGTCA GAAGTAAAGT GGCCGCAAGT TTATCACTCA TGGTTATGGC 9800
 AGCACTGCAT AATTCTCTTA CTGTCATGCC ATCCGTAAGA TGCTTTTCTG 9850
 TGACTGGTGA GTACTCAACC AAGTCATTCT GAGAATAGTG TATGCGGCGA 9900
 CCGAGTTGCT CTTGCCCGCG GTCAATACGG GATAATACCG CGCCACATAG 9950
 CAGAACTTTA AAAGTGCTCA TCATTGGAAA ACGTTCTTCG GGGCGAAAAC 10000
 TCTCAAGGAT CTTACCGCTG TTGAGATCCA GTTCGATGTA ACCCACTCGT 10050
 GCACCCAACT GATCTTCAGC ATCTTTTACT TTCACCAAGC TTTCTGGGTG 10100
 AGCAAAAACA GGAAGGCAAA ATGCCGCAAA AAAGGGAATA AGGGCGACAC 10150
 GGAAATGTTG AATACTCATA CTCTTCCTTT TTCAATATTA TTGAAGCATT 10200
 TATCAGGGTT ATTGTCTCAT GAGCGGATAC ATATTGAAT GTATTTAGAA 10250
 AAATAAACAA ATAGGGGTTC CGCGCACATT TCCCCGAAA GTGCCAC 10297

SEQ ID NO:46 (pTnMod(Oval/ENT tag/P146/PA) - QUAIL)

CTGACGCGCC CTGTAGCGGC GCATTAAGCG CGGCGGGTGT GGTGGTTACG 50
 CGCAGCGTGA CCGCTACACT TGCCAGCGCC CTAGCGCCCG CTCCTTTTCG 100
 TTTCTTCCCT TCCTTTCTCG CCACGTTCGC CGGCATCAGA TTGGCTATTG 150
 GCCATTGCAT ACGTTGTATC CATATCATAA TATGTACATT TATATTGGCT 200
 CATGTCCAAC ATTACCGCCA TGTGACATT GATTATTGAC TAGTTATTAA 250
 TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA TGGAGTTCCG 300
 CGTTACATAA CTTACGGTAA ATGGCCCGCC TGGCTGACCG CCCAACGACC 350
 CCCGCCCAAT GACGTCAATA ATGACGTATG TTCCCATAGT AACGCCAATA 400
 GGGACTTTCC ATTGACGTCA ATGGGTGGAG TATTTACGGT AAAGTGCCCA 450
 CTTGGCAGTA CATCAAGTGT ATCATATGCC AAGTACGCC CCTATTGACG 500
 TCAATGACCG TAAATGGCCC GCCTGGCATT ATGCCAGTA CATGACCTTA 550
 TGGGACTTTC CTAATTGGCA GTACATCTAC GTATTAGTCA TCGCTATTAC 600
 CATGGTGATG CGGTTTTGGC AGTACATCAA TGGGCGTGGA TAGCGGTTTG 650
 ACTCAGGGG ATTTCCAAGT CTCACCCCA TTGACGTCAA TGGGAGTTTG 700
 TTTTGGCACC AAAATCAACG GGAAGTTCCA AAATGTCGTA ACAACTCCGC 750
 CCCATTGACG CAAATGGCG GTAGGCGTGT ACGGTGGGAG GTCTATATAA 800
 GCAGAGCTCG TTAGTGAAC CGTCAGATCG CCTGGAGACG CCATCCACGC 850
 TGTTTTGACC TCCATAGAAG ACACCGGGAC CGATCCAGCC TCCGCGGCCG 900
 GGAACGGTGC ATTGGAACGC GGATTCGCCG TGCCAAGAGT GACGTAAGTA 950
 CCGCTATAG ACTCTATAGG CACACCCCTT TGGCTCTTAT GCATGCTATA 1000
 CTGTTTTTGG CTTGGGGCCT ATACACCCCT GCTTCCCTTAT GCTATAGGTG 1050
 ATGGTATAGC TTAGCCTATA GGTGTGGGTT ATTGACCATT ATTGACCACT 1100
 CCCCTATTGG TGACGATACT TTCCATTACT AATCCATAAC ATGGCTCTTT 1150
 GCCACAAC TAATCTATTG CTATATGCCA ATACTCTGTC CTTAGAGAC 1200
 TGACACGGAC TCTGTATTTT TACAGGATGG GGTCCCATTT ATTATTTACA 1250
 AATTACATA TACAACAACG CCGTCCCCCG TGCCCGCAGT TTTTATTAAA 1300
 CATAGCGTGG GATCTCCAGC CGAATCTCGG GTACGTGTTT CGGACATGGG 1350
 CTCCTTCGCG GTAGCGCGG AGCTTCCACA TCCGAGCCCT GGTCCCATGC 1400
 CTCCAGCGGC TCATGGTGGC TCGGCAGCTC CTTGCTCCTA ACAGTGGAGG 1450
 CCAGACTTAG GCACAGCACA ATGCCACCA CCACCAAGTGT GCCGACAGG 1500
 GCCGTGGCGG TAGGGTATGT GTCTGAAAT GAGCGTGGAG ATTGGGCTCG 1550
 CACGGCTGAC GCAGATGGAA GACTTAAGGC AGCGGCAGAA GAAGATGCAG 1600
 GCAGCTGAGT TGTGTATTTC TGATAAGAGT CAGAGGTAAC TCCCGTTGCG 1650
 GTGCTGTTAA CCGTGGAGGG CAGTGTAGTC TGAGCAGTAC TCGTTGCTGC 1700

	CGCGCGCGCC	ACCAGACATA	ATAGCTGACA	GACTAACAGA	CTGTTCCITT	1750
	CCATGGGTCT	TTTCTGCAGT	CACCGTCGGA	CCATGTGTGA	ACTTGATATT	1800
	TTACATGATT	CTCTTTACCA	ATTCTGCCCC	GAATTACACT	TAAAACGACT	1850
5	CAACAGCTTA	ACGTTGGCTT	GCCACGCATT	ACTTGACTGT	AAAACTCTCA	1900
	CTCTTACCGA	ACTTGGCCGT	AACCTGCCAA	CCAAAGCGAG	AACAAAACAT	1950
	AACATCAAAC	GAATCGACCG	ATTGTTAGGT	AATCGTCACC	TCCACAAGA	2000
	GCGACTCGCT	GTATACCGTT	GGCATGCTAG	CTTTATCTGT	TCGGGAATAC	2050
	GATGCCCATT	GTACTTGTGG	ACTGGTCTGA	TATTCGTGAG	CAAAAACGAC	2100
	TTATGGTATT	GCGAGCTTCA	GTCGCACTAC	ACGGTCGTTT	TGTTACTCTT	2150
10	TATGAGAAAG	CGTTCCCGCT	TTCAGAGCAA	TGTTCAAAGA	AAGCTCATGA	2200
	CCAAATTTCTA	GCCGACCTTG	CGAGCATTCT	ACCGAGTAAC	ACCACACCGC	2250
	TCATTGTCTAG	TGATGCTGGC	TTTAAAGTGC	CATGGTATAA	ATCCGTGAG	2300
	AAGCTGGGTT	GGTACTGGTT	AAGTCGAGTA	AGAGGAAAAG	TACAATATGC	2350
	AGACCTAGGA	GCGGAAAAC	GGAAACCTAT	CAGCAACTTA	CATGATATGT	2400
15	CATCTAGTCA	CTCAAAGACT	TTAGGCTATA	AGAGGCTGAC	TAAAAGCAAT	2450
	CCAATCTCAT	GCCAAATTTCT	ATTGTATAAA	TCTCGCTCTA	AAGGCCGAAA	2500
	AAATCAGCGC	TCGACACGGA	CTCATTGTCA	CCACCCGTCA	CCTAAAATCT	2550
	ACTCAGCGTC	GGCAAAGGAG	CCATGGGTTC	TAGCAACTAA	CTTACCTGTT	2600
	GAAATTCGAA	CACCCAAACA	ACTTGTTAAT	ATCTATTCTGA	AGCGAATGCA	2650
	GATTGAAGAA	ACCTTCCGAG	ACTTGAAGAG	TCCTGCCTAC	GGACTAGGCC	2700
20	TACGCCATAG	CCGAACGAGC	AGCTCAGAGC	GTTTGTGATAT	CATGCTGCTA	2750
	ATCGCCCTGA	TGCTTCAACT	AACATGTTGG	CTTGGCGGCG	TTTATGCTCA	2800
	GAAACAAGGT	TGGGACAAGC	ACTTCCAGGC	TAACACAGTC	AGAAATCGAA	2850
	ACGTACTCTC	AACAGTTCTC	TTAGGCATGG	AAGTTTTCGG	GCATTCTGGC	2900
	TACACAATAA	CAAGGGAAGA	CTTACTCTGT	GCTGCAACCC	TACTAGCTCA	2950
25	AAATTTATTC	ACACATGGTT	ACGCTTTGGG	GAAATTATGA	TAATGATCCA	3000
	GATCACTTCT	GGCTAATAAA	AGATCAGAGC	TCTAGAGATC	TGTGTGTTGG	3050
	TTTTTTGTGG	ATCTGCTGTG	CCTTCTAGTT	GCCAGCCATC	TGTTGTTTGC	3100
	CCCTCCCCCG	TGCCTTCTCT	GACCCCTGGAA	GGTGCCACTC	CCACTGTCTT	3150
	TTCTTAATAA	AATGAGGAAA	TTGCATCGCA	TTGTCTGAGT	AGGTGTCATT	3200
	CTATTCTGGG	GGGTGGGGTG	GGGCAGCACA	GCAAGGGGGA	GGATTGGGAA	3250
30	GACAAATAGCA	GGCATGCTGG	GGATGCGGTG	GGCTCTATGG	GTACCTCTCT	3300
	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CTCTCGGTAC	CTCTGTCTCT	3350
	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CGGTACCAGG	TGCTGAAGAA	3400
	TTGACCCGGT	GACCAAAGGT	GCCTTTTATC	ATCACTTTAA	AAATAAAAAA	3450
	CAATTACTCA	GTGCCGTGTTA	TAAGCAGCAA	TTAATTATGA	TTGATGCCTA	3500
35	CATCACAACA	AAAACGTGAT	TAACAAATGG	TTGGTCTGCC	TTAGAAAGTA	3550
	TATTTGAACA	TTATCTTGAT	TATATTATTG	ATAATAATAA	AAACCTTATC	3600
	CCTATCCAAG	AAGTGATGCC	TATCATTTGG	TGGAATGAAC	TTGAAAAAAA	3650
	TTAGCCTTGA	ATACATTACT	GGTAAGGTAA	ACGCCATTGT	CAGCAAATTG	3700
	ATCCAAGAGA	ACCAACTTAA	AGCTTTCTCT	ACGGAATGTT	AATTCTCGTT	3750
	GACCCTGAGC	ACTGATGAAT	CCCCTAATGA	TTTTGGTAAA	AATCATTAAAG	3800
40	TTAAGGTGGA	TACACATCTT	GTCATATGAT	CCCGGTAATG	TGAGTTAGCT	3850
	CACTCATTAG	GCACCCAGG	CTTTACACTT	TATGCTTCCG	GCTCGTATGT	3900
	TGTGTGGAAT	TGTGAGCGGA	TAACAATTTT	ACACAGGAAA	CAGCTATGAC	3950
	CATGATTACG	CCAAGCGCGC	AATTAACCTT	CACTAAAGGG	AACAAAAGCT	4000
	GGAGCTCCAC	CGCGGTGGCG	GCCGCTCTAG	AACATAGTGA	TCCCCCGGGG	4050
45	AGGTCAGAA	GGTTTCTTTA	CTGTTTGTCA	ATTCTATTAT	TTCAATACAG	4100
	AACAAAAGCT	TCTATAACTG	AAATATATTT	GCTATTGTAT	ATTATGATTG	4150
	TCCTTCGAAC	CATGAACACT	CCTCCAGCTG	AATTTTCAAA	TTCTCTGTTC	4200
	ATCTGCCAGG	CTGGAAGATC	ATGGAAGATC	TCTGAGGAAC	ATTGCAAGTT	4250
	CATACCATAA	ACTCATTGGG	AATTGAGTAT	TATTTTGGCTT	TGAATGGAGC	4300
	TATGTTTTGC	AGTTCCCTCA	GAAGAAAAGC	TTGTTATAAA	GCGTCTACAC	4350
50	CCATCAAAG	ATATATTTAA	ATATTCCAAC	TACAGAAAGA	TTTTGTCTGC	4400
	TCTTCACTCT	GATCTCAGTT	GGTTTCTTCA	CGTACATGCT	TCTTTATTTG	4450
	CCTATTTTGT	CAAGAAAATA	ATAGGTCAAG	TCTGTCTCTC	ACTTATCTCC	4500
	TGCCTAGCAT	GGCTTAGATG	CACGTTGTAC	ATTCAAGAAG	GATCAAATGA	4550
	AACAGACTTC	TGGTCTGTTA	CAACAACCAT	AGTAATAAAC	AGACTAACTA	4600
	ATAATTGCTA	ATTATGTTTT	CCATCTCTAA	GGTTCCCAAC	TTTTTCTGTT	4650
55	TTAAGATCCC	ATTATCTGGT	TGTAAGTGA	GCTCAATGGA	ACATGAACAG	4700
	TATTTCTCAG	TCCTTTCTCC	AGCAATCCTG	ACGGATTAGA	AGAACTGGCA	4750

	GAAAACACTT	TGTTACCCAG	AATTAAAAAC	TAATATTTGC	TCTCCCTTCA	4800
	ATCCAAAATG	GACCTATTGA	AACTAAAATC	TGACCCAATC	CCATTAAATT	4850
5	ATTTCTATGG	CGTCAAAGGT	CAAACCTTTG	AAGGGAACTT	GTGGGTGGGT	4900
	CCCAATTGAG	GCTATATATT	CCCCAGGGCT	CAGCCAGTGG	ATCCATGGGC	4950
	TCCATCGGTG	CAGCAAGCAT	GGAATTTTGT	TTTGATGTAT	TCAAGGAGCT	5000
	CAAAGTCCAC	CATGCCAATG	ACAACATGCT	CTACTCCCCC	TTTGCCATCT	5050
	TGTCAACTCT	GGCCATGGTC	TTCCTAGGTG	CAAAAGACAG	CACCAGGACC	5100
	CAGATAAATA	AGGTGTGTCA	CTTTGATAAA	CTTCCAGGAT	TCGGAGACAG	5150
10	TATTGAAGCT	CAGTGTGGCA	CATCTGTAAA	TGTTCACTCT	TCACTTAGAG	5200
	ACATACTCAA	CCAAATCACC	AAACAAAATG	ATGCTTATTC	GTTGAGCCTT	5250
	GCAGTAGAGC	TTTATGCTCA	AGAGACATAC	ACAGTCGTGC	CGGAATACTT	5300
	GCAATGTGTG	AAGGAACTGT	ATAGAGGAGG	CTTAGAATCC	GTCAACTTTC	5350
	AAACAGCTGC	AGATCAAGCC	AGAGGCCTCA	TCAATGCCTG	GGTAGAAAGT	5400
	CAGACAAACG	GAATTATCAG	AAACATCCTT	CAGCCAAGCT	CCGTGGATTC	5450
15	TCAAACCTGCA	ATGGTCCTGG	TTAATGCCAT	TGCCTTCAAG	GGACTGTGGG	5500
	AGAAAGCATT	TAAGGCTGAA	GACACGCAAA	CAATACCTTT	CAGAGTGACT	5550
	GAGCAAGAAA	GCAAACCTGT	GCAGATGATG	TACCAGATTG	GTTTCATTTAA	5600
	AGTGGCATCA	ATGGCTTCTG	AGAAAATGAA	GATCCTGGAG	CTTCCATTTC	5650
	CCAGTGGAAC	AATGAGCATG	TTGGTGCTGT	TGCCTGATGA	TGTCTCAGGC	5700
20	CTTGAGCAGC	TTGAGAGTAT	AATCAGCTTT	GAAAACTGA	CTGAATGGAC	5750
	CAGTTCTAGT	ATTATGGAAG	AGAGGAAGGT	CAAAGTGATC	TTACCTCGCA	5800
	TGAAGATGGA	GGAGAAATAC	AACCTCACAT	CTCTCTTAAT	GGCTATGGGA	5850
	ATTACTGACC	TGTTGAGCTC	TTCAGCCAAT	CTGTCTGGCA	TCTCTCAGT	5900
	AGGGAGCCTG	AAGATATCTC	AAGCTGTCCA	TGCAGCACAT	GCAGAAATCA	5950
	ATGAAGCGGG	CAGAGATGTG	GTAGGCTCAG	CAGAGGCTGG	AGTGGATGCT	6000
25	ACTGAAGAAT	TTAGGGCTGA	CCATCCATTG	CTCTTCTGTG	TCAAGCACAT	6050
	CGAAACCAAC	GCCATTCTCC	TCTTTGGCAG	ATGTGTTTCT	CCGCGGCCAG	6100
	CAGATGACGC	ACCAGCAGAT	GACGCACCAG	CAGATGACGC	ACCAGCAGAT	6150
	GACGCACCAG	CAGATGACGC	ACCAGCAGAT	GACGCAACAA	CATGTATCCT	6200
	GAAAGGCTCT	TGTGGCTGGA	TCGGCCTGCT	GGATGACGAT	GACAAAAAT	6250
30	ACAAAAAAGC	ACTGAAAAAA	CTGGCAAAAC	TGCTGTAAATG	AGGGCGCCTG	6300
	GATCCAGATC	ACTTCTGGCT	AATAAAAGAT	CAGAGCTCTA	GAGATCTGTG	6350
	TGTTGGTTTT	TTGTGGATCT	GCTGTGCCTT	CTAGTTGCCA	GCCATCTGTT	6400
	GTTTGCCCCC	CCCCCGTGCC	TTCCCTGACC	CTGGAAGGTG	CCACTCCCAC	6450
	TGTCCTTTCC	TAATAAAATG	AGGAAATTGC	ATCGCATTTG	CTGAGTAGGT	6500
	GTCATTCTAT	TCTGGGGGGT	GGGGTGGGGC	AGCACAGCAA	GGGGGAGGAT	6550
35	TGGGAAGACA	ATAGCAGGCA	TGCTGGGGAT	GCGGTGGGCT	CTATGGGTAC	6600
	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CTCTCTCTCT	CGGTACCTCT	6650
	CTCGAGGGGG	GGCCCGGTAC	CCAATTGCGC	CTATAGTGAG	TCGTATTACG	6700
	CGCGCTCACT	GGCCGTCGTT	TTACAACGTC	GTGACTGGGA	AAACCCCTGGC	6750
	GTTATCCCAAC	TTAATCGCCT	TGCAGCACAT	CCCCCTCTCG	CCAGCTGGGC	6800
	TAATAGCGAA	GAGGCCCGCA	CCGATCGCCC	TTCCCAACAG	TTGCGCAGCC	6850
40	TGAATGGCGA	ATGGAAATTG	TAAGCGTTAA	TATTTTGTTA	AAATTGCGGT	6900
	TAAATTTTGT	TTAAATCAGC	TCATTTTTTA	ACCAATAGGC	CGAAATCGGC	6950
	AAAATCCCTT	ATAAATCAAA	AGAATAGACC	GAGATAGGGT	TGAGTGTTGT	7000
	TCCAGTTTGG	AACAAGAGTC	CACATTTAAA	GAACGTGGAC	TCCAACGTCA	7050
	AAGGGCGAAA	AACCGTCTAT	CAGGGCGATG	GCCCACTACT	CCGGGATCAT	7100
	ATGACAAGAT	GTGTATCCAC	CTTAACCTAA	TGATTTTATC	CAAAATCAAT	7150
45	AGGGGATTCA	TCAGTGCTCA	GGGTCAACGA	GAATTAACAT	TCCGTGAGGA	7200
	AAGCTTATGA	TGATGATGTG	CTTAAAAACT	TACTCAATGG	CTGGTTATGC	7250
	ATATCGCAAT	ACATGCGAAA	AACTTAAAAG	AGCTTGCCGA	TAAAAAAGGC	7300
	CAATTTATTG	CTATTTACCG	CGGCTTTTAA	TTGAGCTTGA	AAGATAAATA	7350
	AAATAGATAG	GTTTATTTTG	AAGCTAAATC	TTCTTTATCG	TAAAAAATGC	7400
50	CCTCTTGGGT	TATCAAGAGG	GTCATTATAT	TTGCGGGAAT	AACATCAATT	7450
	GGTGACGAAA	TAACATAAGCA	CTTGCTCTCT	GTTTACTCCC	CTGAGCTTGA	7500
	GGGGTTAACA	TGAAGGTCAT	CGATAGCAGG	ATAATAATAC	AGTAAAACGC	7550
	TAAACCAATA	ATCCAAATCC	AGCCATCCCA	AATTGGTAGT	GAATGATTAT	7600
	AAATAACAGC	AAACAGTAAT	GGGCCAATAA	CACCGGTTGC	ATTGGTAAAG	7650
	CTCACCAATA	ATCCCTGTAA	AGCACCTTGC	TGATGACTCT	TTGTTTGGAT	7700
55	AGACATCACT	CCCTGTAATG	CAGGTAAAGC	GATCCCACCA	CCAGCCAATA	7750
	AAATTAAAC	AGGGAAAAC	AACCAACCTT	CAGATATAAA	CGCTAAAAAG	7800

5 GCAAAATGCAC TACTATCTGC AATAAATCCG AGCAGTACTG CCGTTTTTTC 7850
 GCCCCATTTA GTGGCTATTG TTCTTGCCAC AAAGGCTTGG AATACTGAGT 7900
 GTAAAAGACC AAGACCGGCT AATGAAAAGC CAACCATCAT GCTATTCCAT 7950
 CCAAAACGAT TTTCGGTAAA TAGCACCCAC ACCGTTGCCG GAATTTGGCC 8000
 TATCAATTGC GCTGAAAAAT AAATAATCAA CAAAATGGCA TCGTTTTAAA 8050
 TAAAGTGATG TATACCGAAT TCAGCTTTTG TTCCCTTTAG TGAGGGTTAA 8100
 TTGCGCGCTT GCGTAATCA TGGTCATAGC TGTTTCCTGT GTGAAATTGT 8150
 10 TATCCGCTCA CAATTCCACA CAACATACGA GCCGGAAGCA TAAAGTGTA 8200
 AGCCTGGGGT GCCTAATGAG TGAGCTAACT CACATTAAAT GCGTTGCGCT 8250
 CACTGCCCGC TTTCAGTCG GGAAACCTGT CGTGCCAGCT GCATTAATGA 8300
 ATCGGCCAAC GCGCGGGGAG AGGCGGTTG CGTATTGGGC GCTCTTCCGC 8350
 TTCCTCGCTC ACTGACTCGC TGGCTCGGT CGTTCGGCTG CCGCGAGCGG 8400
 TATCAGCTCA CTCAAAGGCG GTAATACGGT TATCCACAGA ATCAGGGGAT 8450
 15 AACGCAGGAA AGAATCATGT AGCAAAAGG CAGCAAAAG CCAGGAACCG 8500
 TAAAAAGGCC GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC CCCCTGACG 8550
 AGCATCACAA AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCGACAGGA 8600
 CTATAAAGAT ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG TCGCTCTCC 8650
 TGTTCCGACC CTGCGCTTA CCGGATACCT GTCCGCCTTT CTCCTTCCG 8700
 GAAGCGTGCG GCTTTCTCAT AGCTCACGCT GTAGGTATCT CAGTTCGGTG 8750
 20 TAGGTCGTTG GCTCCAAGCT GGGCTGTGTG CACGAACCCC CCGTTCAGCC 8800
 CGACCGCTGC GCCTTATCCG GTAACATATC TCTTGAGTCC AACC CGGTAA 8850
 GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA 8900
 GCGAGGTATG TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAATA 8950
 CCGCTACACT AGAAGGACAG TATTGGTAT CTGCGCTCTG CTGAAGCCAG 9000
 25 TTACCTTCGG AAAAAGAGTT GGTAGCTCTT GATCCGGCAA ACAAAACCACC 9050
 GCTGGTAGCG GTGGTTTTTT TGTGTGCAAG CAGCAGATTA CGCGCAGAAA 9100
 AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC 9150
 AGTGAACGA AAATCAGCT TAAGGGATTT TGGTCATGAG ATTATCAAAA 9200
 AGGATCTTCA CCTAGATCCT TTAAATTA AAATGAAGTT TTAAATCAAT 9250
 30 CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA TGCTTAATCA 9300
 GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTTCAT CATAGTTGCC 9350
 TGACTCCCCG TCGTGTAGAT AACTACGATA CGGGAGGGCT TACCATCTGG 9400
 CCCAGTGCT GCAATGATAC CGCGAGACCC ACGCTCACC GCTCCAGATT 9450
 TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT 9500
 GCAACTTTAT CCGCTCCAT CCAGTCTATT AATTGTTGCC GGAAGCTAG 9550
 35 AGTAAGTAGT TCGCCAGTTA ATAGTTTGCG CAACGTTGTT GCCATTGCTA 9600
 CAGGCATCGT GGTGTCAGCG TCGTCGTTTG GTATGGCTTC ATTACGCTCC 9650
 GGTTCCTAAC GATCAAGGCG AGTTACATGA TCCCCATGT TGTGCAAAA 9700
 AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAGT AAGTTGGCCG 9750
 CAGTGTTATC ACTCATGTT ATGGCAGCAC TGCATAATC TCTTACTGTC 9800
 ATGCCATCCG TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC 9850
 40 ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC CCGGCGTCAA 9900
 TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAGT GCTCATCATT 9950
 GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC CGCTGTTGAG 10000
 ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT 10050
 TFACTTTTAC CAGCGTTTCT GGGTGAGCAA AAACAGGAAG GCAAAATGCC 10100
 45 GCAAAAAGG GAATAAGGGC GACACGGAAA TGTGAAATAC TCATACTCTT 10150
 CCTTTTTCAA TATTATGAA GCATTTATCA GGGTTATTGT CTCATGAGCG 10200
 GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTTCGCGC 10250
 ACATTTCCCC GAAAAGTGCC AC 10272

50

55

SEQ ID NO:47 pTnMCS (CMV-CHOVg-ent-ProInsulin-synPA)

1 ctgacgcgcc ctgtagcggc gcattaagcg cggcgggtgt ggtgggttac cgcagcggtga
 61 cggctacact tgccagcgcc ctagegccc ctcctttcgc tttcttccct tcctttctcg
 121 ccacgttcgc cggcatcaga ttggctattg gccattgcat acgttgatc catatcataa
 181 tatgtacatt tatattggct catgtccaac attaccgcca tgttgacatt gattattgac
 241 tagttattaa tagtaatcaa ttacggggtc attagtcat agcccatata tggagttcgg
 301 cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc ccgcccatt
 361 gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca
 421 atgggtggag tatttacggt aaactgccc cttggcagta catcaagtgt atcatatgcc
 481 aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta

541 catgacctta tgggaacttc ctacttggca gtacatctac gtattagtca tcgctattac
601 catgggtgatg cgggttttggc agtacatcaa tgggcgtgga tagcgggttg actcacgggg
661 atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg
721 ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtagggtgt
781 acgggtggag gtctatataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg
841 ccatccacgc tggtttgacc tccatagaag acacggggac cgtaccagcc tccgcgccg
901 ggaacgggtgc attggaacgc ggattcccg tgccaagagt gacgtaagta ccgctatag
961 actctatagg cacaccctt tggctcttat gcatgctata ctgtttttgg cttggggcct
1021 atacaccccc gcttccttat gctatagggt atgggtatagc ttagcctata ggtgtgggtt
1081 attgaccatt attgaccact cccctatttg tgacgatact ttccattact aatccataac
1141 atgggtcttt gccacaacta tctctatttg ctatatgcca atactctgtc cttcagagac
1201 tgacacggac tctgtatttt tacaggatgg ggtccattt attatttaca aattcacata
1261 tacaacaacg ccgtcccccgc tgcccgcagt ttttattaaa catagcgtgg gatctccacg
1321 cgaatctcgg gtacgtgttc cggacatggg ctcttctcgc gtacggggcg agcttccaca
1381 tccgagccct gggtcccatgc tccagcggc tcattggtcgc tccgagctc cttgctcta
1441 acagtggagg ccagacttag gcacagcaca atgcccacca ccaccagtgt gccgcacaag
1501 gcggtggcgg taggggtatgt gtctgaaaat gagcgtggag attgggctcg cagggtcgac
1561 gcagatggaa gacttaaggc agcggcagaa gaagatgcag gcagctgagt tgttgtattc
1621 tgataagagt cagaggtaac tcccgttgcg gtgctgttaa cgggtggaggg caagtgtagtc
1681 tgagcagtac tegtgtctgc cgcgcgcgcc accagacata atagctgaca gactaacaga
1741 ctgttccctt ccatgggtct tttctgcagt caccgtcgga ccatgtgca actcgatatt
1801 ttacacgact ctctttacca attctgcccc gaattacact taaaaagact caacagctta
1861 acgttggctt gccacgcatt acttgactgt aaaactctca ctctaccga acttggccgt
1921 aacctgcca ccaagcgag aacaaaacat aacatcaaac gaatcgaccg attgttaggt
1981 aatcgctacc tccacaaaga gcgactcgct gtatacgtt ggcattgctag ctttatctgt
2041 tccgggcaata cgtatcccat tgtacttgtt gacttgtctg atattctgta gcaaaaaaga
2101 ctatgggtat tgcgagcttc agtctgacta cagggtcgtt ctgttactct ttatgagaaa
2161 cgttccccc tttcagagca atgttcaaa agactcatg accaatttct agccgacctt
2221 gcgagcattc taccgagtta caccacaccg ctcatgttca gtgatgtcgg ctttaaagt
2281 ccatgtgtata aatccgttga gaagctgggt tgggtactgt taagtctagt aagaggaaaa
2341 gtacaatatg cagacctagg agcggaaaac tggaaacctc tcagcaact acatgatatg
2401 tcatctagtc actcaagac tttaggtat aagaggctga ctaaaagcaa tccaatctca
2461 tgccaaatc tattgtata atctcgtct aaaggccgaa aaaactcagc ctcgacacgg
2521 actcatgtc accaccgtc tactcagcgt cggcaaggga gccatgggtt
2581 ctacgcaacta acttacctgt tgaatttga acacccaaac aacttgtaa tatctattcg
2641 aagcgaatgc agattgaaga aaccttccga gacttgaaaa gtctctgcta cggactaggc
2701 ctacgcccata gccgaacgag cagctcagag cgttttgata tcatgctgct aatcgccctg
2761 atgcttcaac taacatgttg gcttgcgggc gttcatgctc agaaacaagg ttgggacaag
2821 cacttccagg ctaacacagt cagaaatcga aacgtactct caacagttcg cttaggcatg
2881 gaagttttgc ggcattctgg ctacacaata acaagggaag acttactcgt ggcgtgaacc
2941 ctactagctc aaaaattatt caccacatgg taccgtttgg ggaattatg aggggactcg
3001 tctagagcga tccgggatct cgggaaaagc gttggtgacc aaagggtcct ttatcatca
3061 ctttaaaaat aaaaaacaa tactcagtc ctgttataag cagcaattaa ttatgattga
3121 tgccctacat acaacaaaa ctgatttaac aaatgggttg tctgccttag aaagtatatt
3181 tgaacattat cttgattata ttattgataa taataaaaaac cttatcccta tccaagaggt
3241 gatgcctatc attggttggg atgaacttga aaaaaattag ccttgaatac attactggta
3301 aggtaaacgc cattgtcagc aaattgatcc aagagaacca acttaaagct ttcctgacgg
3361 aatgttaatt ctcttgacc ctgagcactg atgaatcccc taatgatttt ggtaaaaatc
3421 attaggttaa ggtggatata catcttgta tatgatcccg gtaatgtgag tttagctact
3481 cattagggac cccaggcttt acacttatg ctccggctc gtatgtgtg tggaaatgtg
3541 agcggataac aatttcacac aggaacacgc tatgaccatg attacgcaa gcgcgcaatt
3601 aaccctcact aaagggaaca aaagctggag ctccaccgcg gtggcgccg ctctagaact
3661 agtggatccc ccgggcatca gattggctat tggccattgc atacgttcta tccatcatc
3721 aatatgtaca ttatatgg ctcatgtcca acattaccgc catgttgaca ttgattattg
3781 actagttatt aatagtaatc aattacgggg tcattagttc atagcccata tatggagttc
3841 cgcgttacat aacttacggg aaatggcccg cctggctgac cgcccaacga cccccgcc
3901 ttgacgtcaa taatgacgta tgttcccata gtaacgcca tagggacttt ccattgacgt
3961 caatgggttg agtatttacg gtaaaactgc cacttggcag tacatcaagt gtatcatatg
4021 ccaagtaacg cccctattga cgtcaatgac ggtaaatggc ccgctggca ttatgccag
4081 tcatgacct tatgggactt tctacttgg cagtacatct acgtattagt catcgctatt
4141 accatgggtga tgcgggtttg gcagtacatc aatgggctg gatagcgggt tgactcacgg
4201 ggatttccaa gtctccaccc cattgacgtc aatgggagtt tgttttggca ccaaaatcaa
4261 cgggactttc caaaatgtgc taacaactcc gccccattga cgcaaatggg cggtaggcgt
4321 gtacgggtgg aggtctatat aagcagagct cgttttagta accgtcagat cgcctggaga
4381 cgccatccac gctgttttga cctccataga agacaccggg accgatccag cctccgccc
4441 cgggaacggg gcattggaac gcggattccc cgtgccaaga gtgacgtaag taccgctat
4501 agactctata ggcacacccc tttggctctt atgcatgcta tactgtttt ggcttggggc
4561 ctatacacc cgccttctt atgctatagg tgatgtata gcttagccta taggtgtggg

5

10

15

20

25

30

35

40

45

50

55

4621 ttattgacca ttattgacca ctccccatt ggtgacgata ctttccatta ctaatccata
 4681 acatggctct ttgccacaac tatctctatt ggctatatgc caatactctg tccctcagag
 4741 actgacacgg actctgtatt ttacaggat ggggtcccat ttattattta caaattcaca
 4801 tatacaacaa cgcgcgtcccc cgtgcccga gtttttatta aacatagcgt gggatctcca
 4861 cgcgaatctc gggtagctgt tccggacatg ggctctctc cggtagcggc ggagcttcca
 4921 catccgagcc ctgggtcccat gccctccagcg gctcatggtc gctcggcagc tcttctgctc
 4981 taacagtggg gggcagactt aggcacagca caatgccac caccaccagt gtgcgcgaca
 5041 aggcgcgtgg ggtagggtat gtgtctgaaa atgagcgtgg agattgggct cgcacggctg
 5101 acgcagatgg aagacttaag gcagcggcag aagaagatgc aggcagctga gttgtgtat
 5161 tctgataaga gtcagaggta actccgttg cgggtgctgt aacgggtggag ggcagtgtag
 5221 tctgagcagt actcgttgc gccgcgcgcg ccaccagaca taatagctga cagactaaca
 5281 gactgttccct ttcctatgggt cttttctgca gtcaccgtcg gatcaatggg ctccatcggt
 5341 gcagcaagca tggaaatttg ttttgatgta ttcaaggagc tcaaatgcca ccatgccaat
 5401 gagaacatct tctactgccc cattgccatc atgtcagctc tagccatggg ataccgtggg
 5461 gcaaaagaca gcaccaggac acaataaat aagggtgttc gctttgataa acttccagga
 5521 ttccggagaca gtattgaagc tcagtgtggc acatctgtaa acgttcactc ttcaactaga
 5581 gatcactcca accaaatcac caaaccaaat gatgtttatt cgttcagcct tgcagctaga
 5641 ctttatgctg aagagagata cccaatcctg ccagaatact tgcagtgtgt gaaggaaactg
 5701 tatagaggag gcttggaaacc tatcaacttt caaacagctg cagatcaagc cagagagctc
 5761 atcaattcct gggtagaaag tcagacaaat ggaattatca gaaatgtcct tcagccaagc
 5821 tccgtggatt ctcaaaactgc aatggttctg gttaatgcca ttgtcttcaa aggacttgg
 5881 gagaaagcat ttaaggatga agacacacaa gcaatgcctt tcagagtgc tgagcaagaa
 5941 agcaaacctg tgcagatgat gtaccagatt ggtttattta gagtggcatc aatggctct
 6001 gagaaaatga agatcctgga gtttccattt gccagtggga caatgagcat gttgggtctg
 6061 ttgcctgatg aagtctcagg ccttgagcag cttgagagat taatcaact tgaanaactg
 6121 actgaatgga ccagtctaa tgttatggaa gagagaagat caaagtgtac ttacctcgca
 6181 tgaagtggg ggaataatc aaacctcacat ctgtcttaat ggctatgggc attactgagc
 6241 tgtttagctc ttcaagccat ctgtctggca tctcctcagc agagagcctg aagatatctc
 6301 aagctgtcca tgcagcacat gcagaaatca atgaagcagg cagagaggtg gttaggtcag
 6361 cagaggctgg agtggatgct gcaagcgtct ctgaagaatt tagggctgac catccattcc
 6421 tcttctgtat caagcacatc gcaaccaaag ccgttctctt cttttggcag atgtgttcc
 6481 cgcggccagc agatgacgca ccagcagatg acgcaacagc agatgacgca ccagcagatg
 6541 acgcaccagc agatgacgca acaacatgta tcttgaagg ctcttctggc tggatcggcc
 6601 tgctggatga cgtatgacaa tttgtgaacc aacacctgtg cggctcacac ctggtggaag
 6661 ctctctacct agtgtgctgg gaaagaggct tcttctacac acccaagacc cgcggggagg
 6721 cagaggacct gcaggtgggg caaggtaggc tgggcggggg cctggtgca ggcagcctgc
 6781 agcccttggc cctgtagggg tccctgcaga agcgtggcat tgtggaacaa tgcgtacca
 6841 gcactctctc cctctaccag ctggagaact actgcaacta gggcgccata agggcgaaat
 6901 atcgcggcgg ctctagacca ggcgcctgga tccagatcac ttctggctaa taaaagatca
 6961 gagctctaga gatctgtgtg ttgggttttt gtggatctgc tgtgtctctt agtggcagc
 7021 catctgttgt ttgccccctc ccgtgacctt ccttgacctt ggaagggtgc actccactg
 7081 tcttttccca ataaatgag gaaattgcat cgcattgtct gagtaggtgt cattctattc
 7141 tgggggggtg ggtggggcag cacagcaagg gggaggattg ggaagacaat agcaggcatg
 7201 ctggggatgc ggtgggtct atgggtacct ctctctctct ctctctctct ctcactctct
 7261 ctctctctcg gtacctctcc tcgagggggg gcccggtacc caattcggcc tatagttagt
 7321 cgtattacgc gcgctcactg gccgtcgttt tacaacgtcg tgactgggaa aacctggcg
 7381 ttacccaact taatcgctt gcagcacatc cccctttcgc cagctggcgt aatagcgaa
 7441 aggcgcgcac cgtatgcctt tcccaacagt tgcgcagcct gaatggcgaa tggaaattgt
 7501 aagcgttaat attttgttaa aattcgcgtt aaatttttgt taaatcagct cattttttaa
 7561 ccaataggcc gaaatcggca aaatccctta taaatcaaaa gaatagaccg agatagggtt
 7621 gagtgtgtgt ccagtttggg acaagagtcc actattaaag aacgtggact ccaacgtcaa
 7681 agggcgaaaa accgtctatc agggcgatgg cccactactc cgggatacata tgacaagatg
 7741 tgtatccacc ttaacttaat gatttttacc aaaatcatca ggggattcat cagtgtcag
 7801 ggtcaacgag aattaacatt ccgtcaggaa agcttatgat gatgatgtgc ttaaaaactt
 7861 actcaatggc tggttatgca tatcgcaata catgcgaaaa acctaaaaga gcttgcgat
 7921 aaaaaaggcc aatttattgc tatttacgcg ggctttttat tgagcttgaa agataaataa
 7981 aatagatagg ttttatttga agctaatact tctttatcgt aaaaaatgcc ctcttgggtt
 8041 atcaagaggg tcattatatt tcgcggaata acatcatttg gtgacgaat aactaagcac
 8101 ttgtctcctg tttactcccc tgagcttgag gggttaacat gaaggctcat gatagcaggg
 8161 taataatata gtaaaacgct aaaccaataa tccaaatcca gccatcccaa attggtagtg
 8221 aatgattata aataacagca aacagtaatg ggccaataac accggttgca ttggtaaggc
 8281 tccaccaata tccctgtaaa gcacctgtct gatgactctt tgtttggata gacatcactc
 8341 cctgtaatgc aggtaaagcg atccccacac cagccaataa aattaaaaca gggaaaacta
 8401 accaaccttc agatataaac gctaaaaagg caaatgcact actatctgca ataagtcgga
 8461 gcagtactgc cgttttttcg cccatttagt ggctattctt cctgccaaa aggcctggaa
 8521 tactgagtgt aaaagaccac gaccgtaat gaaaagccaa ccatcatgct attcatcact
 8581 acgatttctg taatagcacc acaccgtgct ggttggcta tcaatgcgct gaaataataa
 8641 tcaacaaatg gcatcgtaa ataagtgtat tataccgata agcttttgtt ccttttagtg

8701 aggggttaatt ggcgccttgg cgtaatcatg gtcataagctg tttcctgtgt gaaattgtta
 8761 tccgctcaca attccacaca acatacagagc cggaaagcata aagtgtaaag cctgggggtgc
 8821 ctaatgagtg agctaactca cattaattgc gttgcgctca ctgcccgctt tccagtcggg
 8881 aaacctgtcg tgccagctgc attaatgaat cggccaacgc ggggggagag ggggtttgag
 8941 tattggggcgc tcttcgctt cctcgctcac tgactcgctg cgtcggctcg ttcggctgag
 9001 ggcagcggta tcagctcact caaaggcggg aatacgggta tccacagaat caggggataa
 9061 cgcaggaag aacatgtgag caaaaggcca gcaaaaggcc aggaaccgta aaaaaggcgc
 9121 gttgctggcg tttttccata ggctcggccc cctcgacgag catcacaata atcgacgctc
 9181 aagtcagagg tggcgaaacc cgacaggact ataaagatac caggcgtttc cccctggaag
 9241 ctccctcggt cgctctcctg ttccgaccct ggcgcttacc ggatacctgt ccgctttct
 9301 ccttcgggga agcgtggcgc tttctcatag ctacgctgt aggtatctca gttcgggtga
 9361 ggtcgttcgc tccaagctgg gctgtgtgca cgaaccccc gttcagcccg accgctgagc
 9421 cttatccggg aactatcgct ttgagtcaca cccggaaga cagacttat cgcactggc
 9481 agcagccact ggtaacagga ttacgagagc gaggtatgta ggggtgcta cagagttctt
 9541 gaagtggagg cctaactacg gctacactag aaggacagta tttggatctt ggcctctgct
 9601 gaagccagtt accttcggaa aaagagttgg tagctcttga tccggcaaac aaaccaccgc
 9661 tggtagcggg ggtttttttg tttgcaagca gcagattacg cgcagaaaaa aaggatctca
 9721 agaagatcct ttgatctttt ctacgggggtc tgacgctcag tggaaacgaaa actcaggtta
 9781 agggattttg gtcagagat tatcaaaaag gatcttcacc tagatccttt taaattaaaa
 9841 atgaagtttt aaatcaatct aaagtatata tgagtaaaact tggctcgaca gttaccatg
 9901 cttaactcagt gaggaacctc tctcagcgat ctgtctattt cgttcatcca tagttgcctg
 9961 actccccgct gtgtagataa ctacgatacg ggagggttta ccatctggcc ccagtgctgc
 10021 aatgataccg cgagaccac gctcacggc tccagattta tcagcaataa accagccagc
 10081 cggaaaggcc gagcgagaa gtgtcctgc aactttatcc gctccatcc agtctatata
 10141 ttgttgccgg gaagctagag taagttagtt gccagttaat agtttgcca acgttggtg
 10201 cattgctaca ggcacgtgg tgtcacgctc gtcgtttggg atggcttcat tcagtcagg
 10261 ttcccaacga tcaaggcgag ttacatgac ccccatgtt tgcaaaaaag cgttagctc
 10321 cttcggctct cagatcggtg tcagaagtaa gttggccgca gtgttatcac tcatggttat
 10381 ggcagcactg cataattctc ttactgtcat gccatccgta agatgctttt ctgtgactgg
 10441 tgagtactca accaagtcac tctgagaata gtgtatgcgg cgaccgagtt gctcttgccc
 10501 ggcgtcaata cgggataata ccgcgccaca tagcagaact ttaaaagtgc tcatattgg
 10561 aaaaagttct tcggggcgaa aactctcaag gatcttaccg ctgttgagat ccagttcgat
 10621 gtaaccact cgtgcaccca actgatcttc agcatctttt actttacca gcgtttctgg
 10681 gtgagcaaaa acagggaagg aaaaatgcgc aaaaaaggga ataaggcgca cagggaatg
 10741 ttgaatactc ataactcttc tttttcaata ttattgaagc atttatcagg gttattgtct
 10801 catgagcggg tacatatttg aatgtattta gaaaaataaa caaatagggg ttcgcgcac
 10861 atttcccgga aaagtggcac

35 SEQ ID NO:48 (cecropin prepro)
 AAT TTC TCA AGG ATA TTT
 TTC TTC GTG TTC GCT TTG
 GTT CTG GCT TTG TCA ACA
 GTT TCG GCT GCG CCA GAG
 CCG AAA

40
 45 SEQ ID NO:49 (cecropin prepro extended)
 AAT TTC TCA AGG ATA TTT
 TTC TTC GTG TTC GCT TTG
 GTT CTG GCT TTG TCA ACA
 GTT TCG GCT GCG CCA GAG
 CCG AAA TGG AAA GTC TTC
 AAG

50
 SEQ ID NO:50 (cecropin pro)
 GCG CCA GAG CCG AAA

55
 SEQ ID NO:51 (cecropin pro extended)
 GCG CCA GAG CCG AAA TGG AAA GTC TTC AAG

SEQ ID NO:52 (a Kozak sequence)
 ACCATGT

Claims**1. Method of producing proteins, polypeptides or peptides comprising**

5 (i) administering a composition to an oviduct or an ovary of a bird, wherein the composition comprises a transposon-based vector which comprises:

- a) a transposase gene operably linked to a first promoter, the transposase gene encoding for a transposase; and
 10 b) one or more genes of interest operably-linked to one or more additional promoters; wherein the one or more genes of interest and their operably-linked promoters are flanked by transposase insertion sequences recognized by the transposase, and wherein the first promoter comprises a modified Kozak sequence comprising ACCATG (SEQ ID NO:1) and

15 (ii) permitting the one or more genes of interest to be expressed into a protein, a polypeptide or a peptide.

2. The method of claim 1, wherein the composition is to be injected into an artery leading to the oviduct or the ovary.

3. The method of claim 1, wherein the composition is to be injected into a lumen of the oviduct.

20 4. The method of claim 1, wherein the composition further comprises a transfection reagent.

5. The method of claim 1, wherein one to twenty codons at a beginning of the transposase gene are modified by changing a nucleotide at a third base position of the codon to an adenine or thymine without modifying the amino acid encoded by the codon.

25 6. The method of claim 1, wherein the transposon-based vector comprises:

- a) a transposase gene operably-linked to a first promoter and an avian optimized polyA sequence, the transposase gene encoding for a transposase; and
 30 b) one or more genes of interest operably-linked to one or more additional promoters;
 c) wherein the one or more genes of interest and their operably-linked promoters are flanked by transposase insertion sequences recognized by the transposase.

35 7. The method claim 6, wherein the first promoter is a constitutive promoter.

8. The method of claim 6, wherein the first promoter is an oviduct-specific promoter selected from the group consisting of ovalbumin, ovotransferrin, ovomucoid, ovomucin, g2 ovoglobulin, g3 ovoglobulin, ovoflavoprotein, and ovostatin.

40 9. The method of claim 6, wherein the one or more gene of interest is operably-linked to a second promoter.

10. The method of claim 9, wherein the second promoter is an oviduct-specific promoter selected from the group consisting of ovalbumin, ovotransferrin, ovomucoid, ovomucin, g2 ovoglobulin, g3 ovoglobulin, ovoflavoprotein, and ovostatin.

45 11. The method of claim 6, wherein the transposon-based vector further comprises an egg directing sequence or an enhancer operably-linked to the one or more genes of interest.

50 12. The method of any of claims 1 to 11, wherein the animal is a poultry bird.

13. The method of any of claims 1 to 12, wherein the transposase is a Tn10 transposase.

Patentansprüche

55 1. Verfahren zur Herstellung von Proteinen, Polypeptiden oder Peptiden umfassend

(i) Verabreichen einer Zusammensetzung an ein Ovidukt oder ein Ovar eines Vogels, wobei die Zusammen-

setzung einen Transposon-basierten Vektor umfasst, welcher umfasst:

- a) ein Transposase-Gen operativ verknüpft mit einem ersten Promotor, wobei das Transposase-Gen für eine Transposase kodiert, und
- b) eines oder mehrere Gene von Interesse operativ verknüpft mit einem oder mehreren zusätzlichen Promotoren, wobei das eine oder die mehreren Gene von Interesse und ihre operativ verknüpften Promotoren flankiert sind von Transposase-Insertionssequenzen, welche von der Transposase erkannt werden und wobei der erste Promotor eine modifizierte Kozak-Sequenz umfassend ACCATG (SEQ ID NO:1) umfasst und

(ii) Zulassen dass das eine oder die mehreren Gene von Interesse zu einem Protein, einem Polypeptid oder einem Peptid exprimiert werden.

2. Das Verfahren nach Anspruch 1, wobei die Zusammensetzung in eine zum Ovidukt oder zum Ovar führende Arterie zu injizieren ist.

3. Das Verfahren nach Anspruch 1, wobei die Zusammensetzung in ein Lumen des Ovidukts zu injizieren ist.

4. Das Verfahren nach Anspruch 1, wobei die Zusammensetzung ferner ein Transfektionsreagens umfasst.

5. Das Verfahren nach Anspruch 1, wobei ein bis zwanzig Kodons am Anfang des Transposase-Gens modifiziert sind durch Verändern eines Nukleotids in einer dritten Basenposition des Kodons zu einem Adenin oder Thymin ohne Verändern der Aminosäuresequenz, welche durch das Kodon kodiert ist.

6. Das Verfahren nach Anspruch 1, wobei der Transposon-basierte Vektor umfasst:

- a) ein Transposase-Gen operativ verknüpft mit einem ersten Promotor und einer optimierten Vogel-PolyA-Sequenz, wobei das Transposase-Gen für eine Transposase kodiert, und
- b) eines oder mehrere Gene von Interesse, operativ verknüpft mit einem oder mehreren zusätzlichen Promotoren;
- c) wobei das eine oder die mehreren Gene von Interesse und ihre operativ verknüpften Promotoren flankiert sind von Transposase-Insertionssequenzen, welche von der Transposase erkannt werden.

7. Das Verfahren nach Anspruch 6, wobei der erste Promotor ein konstitutiver Promotor ist.

8. Das Verfahren nach Anspruch 6, wobei der erste Promotor ein Ovidukt-spezifischer Promotor ist, welcher ausgewählt ist aus der Gruppe bestehend aus Ovalbumin, Ovotransferrin, Ovomuroid, Ovomucin, g2-Ovoglobulin, g3-Ovoglobulin, Ovoflavoprotein und Ovostatin.

9. Das Verfahren nach Anspruch 6, wobei das eine oder die mehreren Gene von Interesse mit einem zweiten Promotor operativ verknüpft sind.

10. Das Verfahren nach Anspruch 9, wobei der zweite Promotor ein Ovidukt-spezifischer Promotor ist, welcher ausgewählt ist aus der Gruppe bestehend aus Ovalbumin, Ovotransferrin, Ovomuroid, Ovomucin, g2-Ovoglobulin, g3-Ovoglobulin, Ovoflavoprotein und Ovostatin.

11. Das Verfahren nach Anspruch 6, wobei der Transposon-basierte Vektor ferner umfasst eine Ei-dirigierende Sequenz oder einen Enhancer operativ verknüpft mit dem einen oder den mehreren Genen von Interesse.

12. Das Verfahren nach einem der Ansprüche 1 bis 11, wobei das Tier ein Federvieh ist.

13. Das Verfahren nach einem der Ansprüche 1 bis 12, wobei die Transposase eine Tn10-Transposase ist.

Revendications

1. Procédé de production de protéines, de polypeptides ou de peptides comprenant

(i) l'administration d'une composition au niveau d'un oviducte ou d'un ovaire d'un oiseau, la composition comprenant un vecteur à base de transposons comprenant

- a) un gène d'une transposase en liaison fonctionnelle avec un premier promoteur, ledit gène d'une transposase codant pour une transposase ; et
- b) au moins un gène d'intérêt en liaison fonctionnelle avec au moins un promoteur supplémentaire ; lesdits au moins un gène d'intérêt et leurs promoteurs en liaison fonctionnelle étant flanqués par des séquences d'insertion de la transposase reconnues par la transposase, et le premier promoteur comprenant une séquence de Kozak modifiée comprenant ACCATG (SEQ ID NO : 1) et

10

(ii) l'expression dudit au moins un gène d'intérêt dans une protéine, un polypeptide ou un peptide.

2. Procédé selon la revendication 1, dans lequel la composition est destinée à être injectée dans une artère conduisant jusqu'à l'oviducte ou ovaire.

15

3. Procédé selon la revendication 1, dans lequel la composition est destinée à être injectée dans la lumière de oviducte.

4. Procédé selon la revendication 1, dans lequel la composition comprend, en outre, un réactif de transfection.

20

5. Procédé selon la revendication 1, dans lequel de un à vingt codons situés au début du gène de la transposase sont modifiés par échange d'un nucléotide au niveau de la position de la troisième base du codon contre une adénine ou une thymine sans modification de l'acide aminé encodé par le codon.

25

6. Procédé selon la revendication 1, dans lequel le vecteur à base de transposons comprend :

30

- a) le gène d'une transposase en liaison fonctionnelle avec un premier promoteur et une séquence polyA aviaire optimisée, ledit gène d'une transposase codant pour une transposase ; et
- b) au moins un gène d'intérêt en liaison fonctionnelle avec au moins un promoteur supplémentaire ;
- c) lesdits au moins un gène d'intérêt et leurs promoteurs en liaison fonctionnelle étant flanqués par des séquences d'insertion d'une transposase reconnues par la transposase.

7. Procédé selon la revendication 6, dans lequel le premier promoteur est un promoteur constitutif.

35

8. Procédé selon la revendication 6, dans lequel le premier promoteur est un promoteur spécifique de l'oviducte choisi dans le groupe constitué de l'ovalbumine, de l'ovotransferrine, de l'ovomucoïde, de l'ovomucine, de l'ovoglobuline g2, de l'ovoglobuline g3, de l'ovo flavoprotéine et de l'ovostatine.

9. Procédé selon la revendication 6, dans lequel l'au moins un gène d'intérêt est en liaison fonctionnelle avec un second promoteur.

40

10. Procédé selon la revendication 9, dans lequel le second promoteur est un promoteur spécifique de l'oviducte choisi dans le groupe constitué de l'ovalbumine, de l'ovotransferrine, de l'ovomucoïde, de l'ovomucine, de l'ovoglobuline g2, de l'ovoglobuline g3, de l'ovo flavoprotéine et de l'ovostatine.

45

11. Procédé selon la revendication 6, dans lequel le vecteur à base de transposons comprend, en outre, une séquence de contrôle de l'oeuf ou un amplificateur en liaison fonctionnelle avec l'au moins un gène d'intérêt.

12. Procédé selon l'une quelconque des revendications 1 à 11, dans lequel l'animal est une volaille.

50

13. Procédé selon l'une quelconque des revendications 1 à 12, dans lequel la transposase est une transposase Tn10.

55

FIGURE 1



FIGURE 2

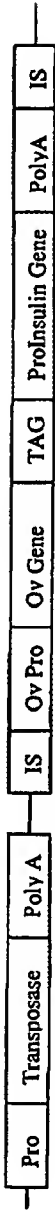


FIGURE 3

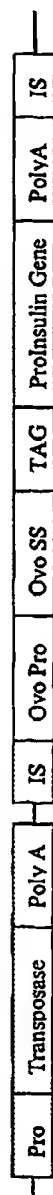
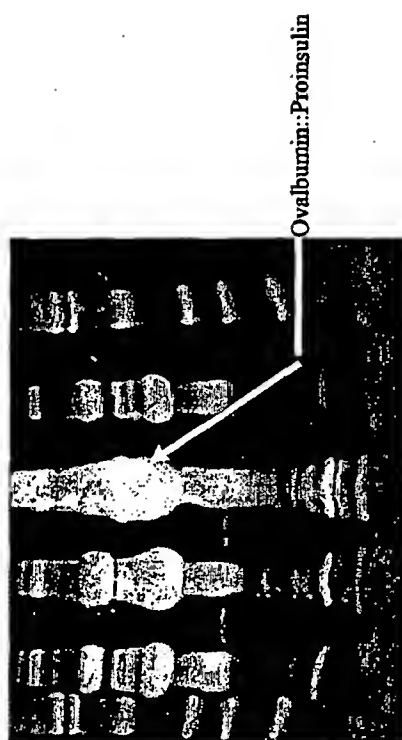


FIGURE 4

IS	Tet, Pro	Ovgen	Pro	Ovotrans	Pro	Ovomucin	IS
----	----------	-------	-----	----------	-----	----------	----

FIGURE 5



REFERENCES CITED IN THE DESCRIPTION

This list of references cited by the applicant is for the reader's convenience only. It does not form part of the European patent document. Even though great care has been taken in compiling the references, errors or omissions cannot be excluded and the EPO disclaims all liability in this regard.

Patent documents cited in the description

- US 5719055 A [0048] [0052]
- US 6291243 B [0054]

Non-patent literature cited in the description

- PIEPER et al. *Diabetes Res. Clin. Pract.*, 1996, S157-S162 [0003]
- KAY, M.A. et al. *Nature Medicine*, 2001, vol. 7, 33-40 [0004] [0005]
- Gene trial to proceed despite fears that therapy could change child's genetic makeup. *The New York Times*, 23 December 2001 [0005]
- *Science, News of the Week*, 04 October 2002 [0005]
- D. LAMPE et al. *Proc. Natl. Acad. Sci. USA*, 1999, vol. 96, 11428-11433 [0054]
- S. FISCHER et al. *Proc. Natl. Acad. Sci. USA*, 2001, vol. 98, 6759-6764 [0054]
- L. ZAGORAIOU et al. *Proc. Natl. Acad. Sci. USA*, 2001, vol. 98, 11474-11478 [0054]
- Mobile DNA, Amer. Soc. Microbiol. 1989 [0054]
- CRONIN, A. et al. *Genes and Development*, 2001, 15 [0061]
- HOPPE, U. C. et al. *Mol. Ther.*, 2000, vol. 1, 159-164 [0061]
- BRASELMANN, S. et al. *Proc. Natl. Acad. Sci.*, 1993, vol. 90, 1657-1661 [0061]
- WANG et al. *Proc. Natl. Acad. Sci.*, 1994, vol. 91, 8180-8184 [0061]
- BELSHAW, P. J. et al. *J. Chem. Biol.*, 1996, vol. 3, 731-738 [0061]
- FAN, L. et al. *Hum. Gene Ther.*, 1999, vol. 10, 2273-2285 [0061]
- SHARIAT, S.F. et al. *Cancer Res.*, 2001, vol. 61, 2562-2571 [0061]
- SPENCER, D.M. *Curr. Biol.*, 1996, vol. 6, 839-847 [0061]
- HOGGATT A.M. et al. *Circ Res.*, 2002, vol. 91 (12), 1151-9 [0062]
- *Biochim Biophys Acta*, 03 January 2003, vol. 1625 (1), 52-63 [0062]
- SIGVARDSSON M. et al. *Mol. Cell Biol.*, 2002, vol. 22 (24), 8539-51 [0062]
- YOSHIMURA I. et al. *J. Urol.*, 2002, vol. 168 (6), 2659-64 [0062]
- ASAKA Y. et al. *Proc. Natl. Acad. Sci.*, 2002, vol. 99 (24), 15456-61 [0062]
- OKINO N. et al. *Biochem. Biophys. Res. Commun.*, 2002, vol. 299 (1), 160-6 [0062]
- GABRIL M.Y. et al. *Gene Ther.*, 2002, vol. 9 (23), 1589-99 [0062]
- KURIKI C. et al. *Biol. Pharm. Bull.*, 2002, vol. 25 (11), 1476-8 [0062]
- STAPLIN W.R. et al. *Blood*, 24 October 2002 [0062]
- BRENNER S. et al. *J. Biol. Chem.*, 18 December 2002 [0062]
- AWADE. Z. *Lebensm. Unters. Forsch.*, 1996, vol. 202, 1-14 [0063]
- Handbook of Fluorescent Probes and Research Products. Molecular Probes, Inc, [0083] [0102]
- GREEN ; WUTS. *Protecting Groups In Organic Synthesis*. John Wiley and Sons, 1991 [0112]
- DITTER et al. *J. Pharm. Sci.*, 1968, vol. 57, 783 [0112]
- DITTER et al. *J. Pharm. Sci.*, 1968, vol. 57, 828 [0112]
- DITTER et al. *J. Pharm. Sci.*, 1969, vol. 58, 557 [0112]
- KING et al. *Biochemistry*, 1987, vol. 26, 2294 [0112]
- LINDBERG et al. *Drug Metabolism and Disposition*, 1989, vol. 17, 311 [0112]
- TUNEK et al. *Biochem. Pharm.*, 1988, vol. 37, 3867 [0112]
- ANDERSON et al. *Arch. Biochem. Biophys.*, 1985, vol. 239, 538 [0112]
- SINGHAL et al. *FASEB J.*, 1987, vol. 1, 220 [0112]
- B. O'MALLEY et al. *EMBO J.*, 1987, vol. 6, 2305-12 [0135]
- A. QIU et al. *Proc. Nat. Acad. Sci. (USA)*, 1994, vol. 91, 4451-4455 [0135]
- D. MONROE et al. *Biochim. Biophys. Acta*, 2000, vol. 1517 (1), 27-32 [0135]
- H. PARK et al. *Biochem.*, 2000, vol. 39, 8537-8545 [0135]
- T. MURAMATSU et al. *Poult. Avian Biol. Rev.*, 1996, vol. 6, 107-123 [0135]
- Egg Science & Technology. Haworth Press, 1995 [0140] [0145]
- Practical Protein Chemistry A Handbook. John Wiley & Sons Ltd, 1986 [0147]
- T. OKA ; RT SCHIMKE. *J. Cell Biol.*, 1969, vol. 41, 816 [0196]

- PALMITER ; CHRISTENSEN ; SCHIMKE. *J Biol. Chem.*, 1970, vol. 245 (4), 833-845 [0196]